

UNIVERSITÉ DU QUÉBEC

MÉMOIRE PRÉSENTÉ À L'UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES

COMME EXIGENCE PARTIELLE
DE LA MAÎTRISE EN MATHÉMATIQUES
ET INFORMATIQUE APPLIQUÉES

PAR
Alexandre Dumas

Nouvelles extensions de la loi exponentielle bivariée de Raftery

Mars 2019

Université du Québec à Trois-Rivières

Service de la bibliothèque

Avertissement

L'auteur de ce mémoire ou de cette thèse a autorisé l'Université du Québec à Trois-Rivières à diffuser, à des fins non lucratives, une copie de son mémoire ou de sa thèse.

Cette diffusion n'entraîne pas une renonciation de la part de l'auteur à ses droits de propriété intellectuelle, incluant le droit d'auteur, sur ce mémoire ou cette thèse. Notamment, la reproduction ou la publication de la totalité ou d'une partie importante de ce mémoire ou de cette thèse requiert son autorisation.

Remerciements

Premièrement, je souhaite adresser mes plus sincères remerciements à toutes les personnes m'ayant apporté leur aide et qui ont contribué à la réussite de mes deux années universitaires ainsi qu'à la réalisation de ce mémoire.

Je tiens à sincèrement remercier Monsieur Mesfioui Mhamed, Professeur titulaire au département de mathématiques et informatiques de l'université du Québec à Trois-Rivières, pour sa disponibilité tout au long de l'élaboration de ce mémoire, ainsi qu'à son aide et le temps qu'il m'a consacré.

Je tiens également remercier mes parents, ma soeur, mes proches ainsi que toute ma famille pour le long soutien tout au long de mes études.

Mon travail de recherche a été financé par des octrois individuels accordés à M.Mhamed Mesfioui, ainsi que par le conseil de recherches en sciences naturelles et en génie du Canada.

Table des matières

Résumé	ii
Table des matières	v
1 Introduction	1
2 Rappels sur les variables et les vecteurs aléatoires continus	3
2.1 Espace échantillonnage	3
2.1.1 Ensemble des résultats	3
2.1.2 Définition d'une probabilité et axiomatique	4
2.2 Définition d'une variable aléatoire continue	4
2.3 Densité d'une variable aléatoire continue	5
2.4 Fonction de répartition d'une variable aléatoire continue	6
2.5 Espérance et variance d'une variable aléatoire continue	7
2.5.1 Espérance d'une variable aléatoire continue	7
2.5.2 Variance d'une variable aléatoire continue	10
2.6 Exemples de variables aléatoires continues	10
2.6.1 Loi exponentielle	11
2.6.2 Loi Gamma	12
2.6.3 Loi normale	13
2.7 Définition d'un vecteur aléatoire continu	14
2.8 Densité d'un vecteur aléatoire continu	15
2.9 Densité marginale d'un vecteur aléatoire continu	16
2.10 Exemples de vecteurs aléatoires continus	17
3 Rappel sur les copules	19
3.1 Théorème de Sklar	19
3.2 Définition des copules	20
3.3 Construction d'une loi multidimensionnelle à travers une copule	21
3.4 Construction d'une copule à travers une distribution multidimensionnelle	21
3.5 Copules de survie	22
3.6 Quelques copules classiques	23

3.6.1	Copule Indépendante	23
3.6.2	Bornes de Fréchet-Hoeffding	24
3.6.3	Copule Archimédienne	24
3.6.4	Simulation de quelques copules classiques	26
3.7	Mesures de concordance : Tau de Kendall et rho de Spearman	32
3.7.1	Tau de Kendall	33
3.7.2	Rho de Spearman	34
3.8	Expression du Tau de Kendall et Rho de Spearman de la copule sous-jacente	35
4	Nouvelle loi exponentielle bidimensionnelle de type Raftery	38
4.1	Modèle exponentiel bidimensionnel de Raftery	38
4.2	Nouveau modèle exponentiel bidimensionnel de type Raftery	39
4.2.1	Structure de corrélation du modèle exponentiel proposée $\mathcal{E}(\theta, \lambda_1, \lambda_2)$	41
4.3	Copule de survie associée à la famille $\mathcal{E}(\theta, \lambda_1, \lambda_2)$	42
4.3.1	Algorithme de simulation de la copule de survie de la famille $\mathcal{E}(\theta, \lambda_1, \lambda_2)$	45
4.4	Mesures de concordances correspondantes à la famille $\mathcal{E}(\theta, \lambda_1, \lambda_2)$	47
4.4.1	Tau de Kendall de la famille $\mathcal{E}(\theta, \lambda_1, \lambda_2)$	47
4.4.2	Rho de Spearman de la famille $\mathcal{E}(\theta, \lambda_1, \lambda_2)$	48
4.5	Étude de l'estimateur des moments du paramètre de dépendance de la copule de Raftery	49
4.5.1	Estimateur du paramètre de dépendance à partir de la corrélation	49
4.5.2	Estimateur du paramètre de dépendance à partir du rho de Spearman	50
4.5.3	Estimateur du paramètre de dépendance à partir du tau de Kendall	50
4.6	Comparaison des estimateurs par simulation	51
4.7	Fonction de survie de la famille exponentielle bivariée $\mathcal{E}(\theta, \lambda_1, \lambda_2)$	53
4.7.1	Algorithme de simulation de la fonction de survie de la famille $\mathcal{E}(\theta, \lambda_1, \lambda_2)$	54
5	Extension du modèle de Raftery au cas de dépendance négative	56
5.1	Modèle exponentiel bidimensionnel négatif de type Raftery	56
5.2	Structure de corrélation de la famille $\mathcal{E}^-(\lambda_1, \lambda_2, \theta)$	58
5.3	Copule de survie	59
5.3.1	Algorithme de simulation de la copule de survie de la famille $\mathcal{E}(\theta, \lambda_1, \lambda_2)$	62
5.4	Rho de Spearman	63
5.5	Étude de l'estimateur des moments du paramètre de dépendance de la copule de Raftery au cas de dépendance négative	64
5.5.1	Estimateur du paramètre de dépendance à partir de la corrélation	64

5.5.2	Estimateur du paramètre de dépendance à partir de la corrélation de Spearman	65
5.6	Comparaison des estimateurs par simulation	66
5.6.1	Algorithme de simulation de la fonction de survie de la famille $\mathcal{E}^-(\theta, \lambda_1, \lambda_2)$	68
6	Conclusion	71
	Bibliographie	72

Chapitre 1

Introduction

Raftery et al (6) (1984) a proposé un modèle exponentiel bidimensionnel construit selon la méthode de choc commun. Cette idée se base sur un principe simple permettant de créer une dépendance entre des composantes exponentielles indépendantes via une même variable aléatoire appelée choc commun. Il importe de mentionner que le modèle exponentiel bivarié de Raftery construit par cette méthode présente deux inconvénients principaux. Le premier désavantage est que cette distribution modélise uniquement les paires de variables aléatoires identiquement distribuées. Le deuxième désavantage est que le modèle proposé par Raftery et al (6) (1984) capte seulement la dépendance positive.

Les objectifs principaux de ce mémoire consistent à établir des extensions du modèle exponentiel bidimensionnel de Raftery permettant d'éviter les inconvénients cités précédemment. Dans un premier temps, nous proposons une extension du modèle de Raftery capable de modéliser des paires aléatoires non identiquement distribuées. Ceci est rendu possible grâce à la méthode de chocs comonotones suggérée récemment par Genest, Mesfioui et Shoulz (2018). En particulier, nous avons établi, pour ce nouveau modèle, une procédure d'estimation du paramètre de dépendance basée sur la méthode des moments. Dans un deuxième temps, nous avons construit un deuxième modèle spécifique pour la dépendance négative en utilisant le principe de choc contre-comonotone basé sur la borne de Fréchet inférieure. Ce nouveau modèle a donné naissance à une nouvelle copule, qui peut-être utilisée en pratique pour modéliser la dépendance négative. Quelques propriétés de cette nouvelle distribution ont été discutés. En particulier, des simulations ont été effectuées afin de comparer l'efficacité de chaque approche.

Ce mémoire est structuré comme suit. Le second chapitre est divisé en deux parties,

soit les variables aléatoires continues ainsi que les vecteurs aléatoires continus. Nous rappellerons quelques définitions ainsi que les principales notions sur les variables aléatoires et les vecteurs aléatoires. Pour ce faire, nous présenterons quelques exemples de variables et de vecteurs aléatoires.

Le troisième chapitre fait l'objet de l'étude général des copules. Nous rappellerons le théorème de Sklar, ainsi que ses utilisations. Par la suite, nous étudierons la construction d'une loi multidimensionnelle à travers une copule, puis la construction d'une copule à travers une distribution multidimensionnelle à l'aide d'exemples ainsi que de simulations. Également, nous conclurons par l'étude des mesures de concordance, soit le tau de Kendall et le rho de Spearman, pour ensuite exprimer ses mesures de concordances en fonction de la copule sous-jacente.

Le quatrième chapitre représente la première contribution de ce travail. Il s'agit de l'introduction d'une loi exponentielle bidimensionnelle de type Raftery plus générale capable de représenter des variables aléatoires non échangeables. Ce qui permet de remédier au problème du modèle de Raftery classique qui se limite à la modélisation de paires de variables aléatoires identiquement distribuées. Plusieurs propriétés intéressantes de ce nouveau modèle seront discutées dans ce chapitre.

La seconde contribution importante de ce travail se situe au chapitre cinq. Dans cette partie, nous proposons une nouvelle extension de la loi exponentielle bidimensionnelle de Raftery au cas de dépendance négative. Ceci utilise une idée récente permettant de lier des composantes indépendantes par des variables aléatoires parfaitement négativement dépendantes. Nous allons, entre autres, présenter la structure de corrélation du nouveau modèle. Nous allons extraire à partir de ce nouveau modèle une copule originale. Quelques paramètres de concordances associés à cette copule seront discutés. L'estimation du paramètre de dépendance sera aussi présenté. Un algorithme de simulation de cette copule sera établi. Ceci permettra de produire quelques simulations qui seront utilisées afin d'apprécier la qualité des estimateurs proposés.

Chapitre 2

Rappels sur les variables et les vecteurs aléatoires continus

Dans ce chapitre, il sera mentionné d'un rappel général sur les variables et vecteurs aléatoires continus. En débutant par l'axiomatique et la définition d'une probabilité, on établira quelques grandes notions élémentaires mais nécessaires pour introduire les notions futures. En première partie, nous allons définir les variables aléatoires continues et étudier la densité, la fonction de répartition, l'espérance et la variance de plusieurs variables aléatoires connues. En second lieu, il sera mentionné des vecteurs aléatoires continus, le calcul de leur densité ainsi que leur densité marginale. La plupart des résultats présentés ici sont repris du livre de Ross (8).

2.1 Espace échantillonnage

2.1.1 Ensemble des résultats

Supposons une expérience où l'on désire étudier un phénomène aléatoire. Par exemple, le nombre de personnes qui assistent à un spectacle, ou bien le temps pour compléter un marathon de trente kilomètres. Il faut d'abord connaître l'ensemble des valeurs possibles. Cet ensemble de valeurs possibles sera dénommé Ω . L'ensemble Ω peut être soit dénombrable fini, dénombrable infini ou non dénombrable.

2.1.2 Définition d'une probabilité et axiomatique

Avant d'introduire les variables aléatoires, nous allons nous inspirer de la définition de Ross (8). Nous définirons la notion de probabilité ainsi que l'axiomatique des probabilités. En mathématiques, une probabilité est une fonction réelle notée P , qui à chaque événement A dont on ne peut prédire l'issue, fait correspondre un réel d'un espace Ω noté :

$$P: A \rightarrow P(A).$$

Définition 2.1.1. *On suppose qu'une expérience d'un ensemble fondamental S est répétée plusieurs fois exactement dans les mêmes conditions. Soit E un événement favorable, c'est-à-dire que l'on veut voir se produire, on définit $n(E)$ le nombre de fois que E s'est produit lors des n répétitions. Alors $P(E)$ la probabilité que l'évènement E se produise est*

$$P(E) = \lim_{n \rightarrow \infty} \frac{n(E)}{n}.$$

Admettons qu'un nombre $P(E)$ existe et respecte les trois axiomes suivants.

Axiome 2.1.1. *Soit un ensemble fondamental S et un événement E quelconque.*

$$0 \leq P(E) \leq 1,$$

$$P(S) = 1.$$

Axiome 2.1.2. *Pour chaque paire d'évènements mutuellement exclusifs E_1, E_2, E_3, \dots*

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i).$$

$P(E)$ est la probabilité que l'évènement E se réalise.

2.2 Définition d'une variable aléatoire continue

Il existe des variables dont l'ensemble des états possibles est infini non dénombrable. Prenons par exemple, la durée de vie d'un téléphone portable ou bien la quantité annuelle de précipitations dans une ville. Utilisons X comme variable. On définira X comme une variable aléatoire continue de cette façon :

Définition 2.2.1. Une variable aléatoire X est continue s'il existe une fonction f non négative définie pour tout $x \in \mathbb{R}$ tel que pour tout ensemble A de nombres réels la propriété suivante est vérifiée.

$$P(X \in A) = \int_A f(x) dx.$$

Donc, une variable aléatoire continue est une fonction qui exprime le résultat d'une expérience à un nombre réel. On peut noter cette fonction de cette façon :

$$X : \Omega \rightarrow \mathbb{R}.$$

Par exemple, la durée de vie d'une marque d'ordinateur est une variable aléatoire continue. On peut associer une probabilité que l'ordinateur ne soit pas défectueux pendant un intervalle de temps. Le temps est une variable continue.

2.3 Densité d'une variable aléatoire continue

La fonction f , mentionnée précédemment dans (2.2.1), sera dénommée densité de probabilité de la variable aléatoire X . En utilisant la définition d'une variable aléatoire continue, on peut évaluer la probabilité que X prenne une valeur dans A . Il suffit d'intégrer la densité de probabilité sur A . Puisque X prend obligatoirement une valeur, on peut déduire une contrainte sur la fonction f :

$$P\{X \in (-\infty, \infty)\} = \int_{-\infty}^{\infty} f(x) dx = 1.$$

Alors, pour un problème spécifique de probabilité, on peut retrouver la probabilité que X soit dans l'intervalle $A = [a, b]$. En utilisant (2.2.1), on obtient

$$P(a \leq x \leq b) = \int_a^b f(x) dx \leq 1.$$

Cette intégrale représente l'aire sous la courbe dans l'intervalle $[a, b]$. Alors la probabilité d'être dans $[a, b]$ est le rapport entre l'aire sous la courbe dans $[a, b]$ sur l'aire sous la courbe dans $[-\infty, \infty]$. En posant $[a, b] = (-\infty, \infty)$ on obtient le résultat précédent, et lorsque $[a, b] \subseteq (-\infty, \infty)$ on peut affirmer que la probabilité sera inférieure à 1. Si l'on veut trouver la probabilité qu'une variable aléatoire continue prenne une certaine valeur, on utilise (2.3) et en posant $a=b$ on trouve

$$P(X = a) = \int_a^a f(x) dx = 0.$$

Puisque la probabilité qu'une variable aléatoire ait une valeur fixe est nulle, on peut alors établir que

$$P(x < a) = P(x \leq a) = F(a) = \int_{-\infty}^a f(x) dx.$$

Définition 2.3.1. Soit f une fonction continue non-négative. Alors f est une densité si :

$$\int_{-\infty}^{\infty} f(x) dx = 1.$$

Exemple 2.3.1. Supposons que X est une variable aléatoire continue dont la densité est

$$f(x) = \begin{cases} kx^2 & \text{si } x \in [0, 1], \\ 0 & \text{sinon.} \end{cases}$$

Pour trouver la valeur de k , puisque f respecte la condition pour être une densité, c'est-à-dire $\int_{-\infty}^{\infty} f(x) dx = 1$. Par la densité définie précédemment, on trouve que

$$1 = \int_0^1 kx^2 dx = \frac{kx^3}{3} \Big|_0^1 = \frac{k}{3} \Leftrightarrow k = 3.$$

Alors f est bien une densité et est définie ainsi $f(x) = \begin{cases} 3x^2 & \text{si } x \in [0, 1], \\ 0 & \text{sinon.} \end{cases}$. Trouvons

maintenant $P\left(x > \frac{1}{2}\right)$

$$P\left(x > \frac{1}{2}\right) = \int_{1/2}^1 3x^2 dx = x^3 \Big|_{1/2}^1 = 1 - \left(\frac{1}{2}\right)^3 = \frac{7}{8}.$$

Par la densité de f définie précédemment, $P\left(x > \frac{1}{2}\right) = \frac{7}{8}$.

2.4 Fonction de répartition d'une variable aléatoire continue

Il existe une fonction, exprimant l'aire sous la courbe accumulée de la densité sur l'intervalle $(-\infty, a]$. Cette fonction est dénommée fonction de répartition. On la notera

F . Dans la plupart du temps, la fonction de répartition d'une variable aléatoire continue nous permet d'évaluer en chaque point $P(x < a)$.

Définition 2.4.1. Soit F une fonction et f la densité d'une variable aléatoire continue X . F est la fonction de répartition de X si

$$F(a) = P(X \in (-\infty, a]) = \int_{-\infty}^a f(x) dx.$$

En dérivant la fonction de répartition, on obtient

$$\frac{d}{da}(F(a)) = f(a).$$

Donc la densité d'une variable aléatoire est la dérivée de la fonction de répartition.

Exemple 2.4.1. En utilisant $F(a) = \int_{-\infty}^a f(x) dx$ la fonction de répartition associée à la densité de l'exemple (2.3.1) :

$$F(a) = \begin{cases} 0 & \text{si } x \leq 0, \\ x^3 & \text{si } 0 < x < 1, \\ 0 & \text{sinon.} \end{cases}.$$

On peut donc calculer facilement $P\left(x > \frac{1}{2}\right) = 1 - F\left(\frac{1}{2}\right) = \frac{7}{8}$.

Alors la densité et la fonction de répartition d'une variable aléatoire continue sont reliées l'une à l'autre.

2.5 Espérance et variance d'une variable aléatoire continue

L'espérance et la variance d'une variable aléatoire sont définies de façons différentes qu'une variable aléatoire continue. Par contre, l'approche est basée sur le cas continu pour l'espérance et la variance.

2.5.1 Espérance d'une variable aléatoire continue

Dans le cas d'une variable aléatoire discrète X , l'espérance est définie par

$$E(X) = \sum_x xP(x = X).$$

Par contre, on sait que $P(X = a) = 0$, donc on obtiendrait toujours une valeur nulle. Alors en utilisant la fonction de densité $f(x)$, on doit utiliser la même approche que l'intégrale de Riemann, c'est-à-dire :

$$f(x)dx \approx P(x \leq X \leq x + dx).$$

En utilisant l'approximation précédente, on peut alors utiliser l'intégrale pour correctement définir l'espérance de X ainsi :

Définition 2.5.1. Soit X une variable aléatoire continue et f la densité de X . L'espérance de la variable X est :

$$E(X) = \int_{-\infty}^{\infty} xf(x) dx.$$

Exemple 2.5.1. Soit X une variable aléatoire continue telle que sa fonction de densité est $f(x) = \begin{cases} 3x^2 & \text{si } x \in [0, 1], \\ 0 & \text{sinon.} \end{cases}$ Puisque f est définie sur l'intervalle $[0, 1]$, on a que :

$$E(X) = \int_0^1 x(3x^2) dx = \frac{3x^4}{4} \Big|_0^1 = \frac{3}{4}.$$

Par la suite pour la définition de la variance, il sera pratique de calculer l'espérance d'une fonction d'une variable aléatoire continue. Par exemple, on peut chercher $E(X^2)$. Pour ce faire, il faudra utiliser le théorème suivant :

Théorème 2.5.1. Si X est une variable aléatoire de densité $f(x)$, alors pour toute fonction réelle g non-négative :

$$E(g(X)) = \int_{-\infty}^{\infty} f(x)g(x)dx.$$

Avant de faire la démonstration, il faudra utiliser un autre théorème ci-dessous :

Théorème 2.5.2. Soit Y une variable aléatoire non négative,

$$E(Y) = \int_0^{\infty} P(Y > y)dy.$$

Démonstration. Puisque Y est une variable aléatoire continue de densité f_Y , on a

$$P(Y > y)dy = \int_y^{\infty} f_Y(x)dx.$$

En utilisant ce fait, on a :

$$\int_0^{\infty} P(Y > y)dy = \int_0^{\infty} \int_y^{\infty} f_Y(x)dx dy.$$

En changeant l'ordre d'intégration de la dernière équation, on a

$$\int_0^\infty P(Y > y)dy = \int_0^\infty \left(\int_0^x dy \right) f_Y(x)dx = \int_0^\infty x f_Y(x)dx = E(Y).$$

□

Maintenant, démontrons le théorème (2.5.1) en utilisant le précédent.

Démonstration. Par le théorème (2.5.2) Ross (8) a montré que

$$\begin{aligned} E(g(X)) &= \int_0^\infty P(Y > y)dy \\ &= \int_0^\infty \int_{g(x) > y} f(x)dx dy \\ &= \int_{g(x) > 0} \int_0^{g(x)} dy f(x)dx \\ &= \int_{g(x) > 0} g(x) f(x)dx. \end{aligned}$$

ce qui conclut la démonstration.

□

Avant de définir la variance d'une variable aléatoire continue, l'exemple suivant utilisant le théorème (2.5.1) permettra de calculer $E(X^2)$. Par le théorème (2.5.1), on a :

$$E(X^2) = \int_{-\infty}^\infty x^2 f(x) dx.$$

Exemple 2.5.2. Soit X une variable aléatoire continue tel que sa fonction de densité est $f(x) = \begin{cases} 3x^2 & \text{si } x \in [0, 1], \\ 0 & \text{sinon.} \end{cases}$ Puisque f est définie sur l'intervalle $[0, 1]$, on a que :

$$E(X^2) = \int_0^1 x^2 (3x^2) dx = \frac{3x^5}{5} \Big|_0^1 = \frac{3}{5}.$$

Pour l'instant, on a que $E(X) = \frac{3}{4}$ et $E(X^2) = \frac{3}{5}$.

2.5.2 Variance d'une variable aléatoire continue

La variance d'une variable aléatoire continue sert à caractériser la dispersion d'une distribution. Elle est définie de la même façon que dans le cas discret c'est-à-dire :

$$\begin{aligned}
 \text{Var}(X) &= E((X - \mu)^2) \\
 &= E(X^2 - 2\mu X + \mu^2) \\
 &= E(X^2) + E(-2\mu X) + E(\mu^2) \\
 &= E(X^2) - 2\mu E(X) + E(\mu^2) \\
 &= E(X^2) - E(X)^2.
 \end{aligned}$$

Définition 2.5.2. Soit X une variable aléatoire continue. Alors la variance de X est :

$$\text{Var}(X) = E(X^2) - E(X)^2.$$

Alors pour déterminer la variance, on utilise le théorème 2.5.1 pour calculer :

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) dx.$$

Exemple 2.5.3. Reprenons la même variable aléatoire X que dans les exemples précédant comme densité : $f(x) = \begin{cases} 3x^2 & \text{si } x \in [0, 1], \\ 0 & \text{sinon.} \end{cases}$ Calculons la variance.

Puisque $E(X) = \frac{3}{4}$ et $E(X^2) = \frac{3}{5}$. Alors la variance est :

$$\begin{aligned}
 \text{Var}(X) &= E(X^2) - E(X)^2 \\
 &= \frac{3}{5} - \left(\frac{3}{4}\right)^2 \\
 &= \frac{3}{80}.
 \end{aligned}$$

2.6 Exemples de variables aléatoires continues

Dans cette section, on étudiera le comportement de certaines variables aléatoires continues classiques. Pour chacune, on s'intéressera à leur densité, leur fonction de répartition, leur espérance et leur variance. On étudiera la loi exponentielle, la loi Gamma ainsi que la loi normale.

2.6.1 Loi exponentielle

Définition 2.6.1. Une variable aléatoire est une variable aléatoire exponentielle si la densité est définie pour un λ positif :

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x \geq 0, \\ 0 & \text{sinon} \end{cases}.$$

La fonction de répartition associée à cette densité est :

$$\begin{aligned} F(a) &= P(X \leq a) \\ &= \int_0^a \lambda e^{-\lambda x} dx \\ &= -e^{-\lambda x} \Big|_0^a \\ &= 1 - e^{-\lambda x}, a \geq 0. \end{aligned}$$

On peut vérifier que la fonction de répartition est bien définie car $\lim_{a \rightarrow \infty} F(a) = 1$. Trouvons l'espérance et la variance de la variable aléatoire exponentielle.

$$\begin{aligned} E(X) &= \int_0^\infty x \lambda e^{-\lambda x} dx \\ &= -\frac{1}{\lambda} e^{-\lambda x} (\lambda x + 1) \Big|_0^\infty \\ &= \frac{1}{\lambda}. \end{aligned}$$

Maintenant pour la variance, on a :

$$\begin{aligned} E(X^2) &= \int_0^\infty x^2 \lambda e^{-\lambda x} dx \\ &= -\frac{1}{\lambda^2} e^{-\lambda x} (\lambda^2 x^2 + 2\lambda x + 2) \Big|_0^\infty \\ &= \frac{2}{\lambda^2} \\ \text{Var}(X) &= E(X^2) - E(X)^2 \\ &= \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 \\ &= \frac{1}{\lambda^2}. \end{aligned}$$

La loi exponentielle est relativement présente dans les situations lorsqu'on attend qu'un évènement se passe dans le temps. Par exemple, la durée avant le bris d'une pièce d'ordinateur ou bien l'attente avant de recevoir un appel téléphonique. En utilisant la fonction de répartition de la loi exponentielle, on peut déterminer $P(X \leq a)$ ou même $P(X \geq a)$. Il suffit de constater que $P(X \leq a) = F(a) = \int_0^a \lambda e^{-\lambda x} dx$.

Exemple 2.6.1. Soit X une variable aléatoire exponentielle de paramètre $\lambda = 1/2$. Trouvons $P(X \leq 2)$.

$$P(X \leq 2) = F(2) = 1 - e^{-1} \approx 0.632.$$

2.6.2 Loi Gamma

Pour pouvoir introduire la loi Gamma, il faut d'abord définir la fonction gamma.

Définition 2.6.2. Soit $\Gamma(s)$ appelée, fonction gamma, est que $\forall x > 0$ on a :

$$\Gamma(s) = \int_0^{\infty} e^{-x} x^{s-1} dx.$$

En intégrant par partie successivement $\Gamma(s)$, on a :

$$\Gamma(s) = (s-1)\Gamma(s-1).$$

Exemple 2.6.2. Calculons $\Gamma(1)$.

$$\begin{aligned} \Gamma(1) &= \int_0^{\infty} e^{-x} x^{1-1} dx \\ &= \int_0^{\infty} e^{-x} dx \\ &= 1. \end{aligned}$$

Alors du fait que $\Gamma(1) = 1$, on a que si n est un entier, alors $\Gamma(n) = (n-1)!$.

Définition 2.6.3. On dira qu'une variable aléatoire continue X suit une loi gamma de paramètres (s, λ) , tel que $\lambda > 0$ et $s > 0$ si sa densité est :

$$f(x) = \begin{cases} \frac{\lambda e^{-\lambda x} (\lambda x)^{s-1}}{\Gamma(s)} & \text{si } x \geq 0, \\ 0 & \text{sinon.} \end{cases}$$

On peut remarquer que la loi gamma de paramètre $(1, \lambda)$ est tout simplement une variable aléatoire exponentielle. Trouvons l'espérance et la variance de la loi gamma. En utilisant le fait que $\Gamma(n) = (n-1)\Gamma(n-1)$, $\forall n \in \mathbb{N}$ on a :

$$\begin{aligned} E(X) &= \frac{1}{\Gamma(s)} \int_0^{\infty} \lambda x e^{-\lambda x} (\lambda x)^{s-1} dx \\ &= \frac{1}{\lambda \Gamma(s)} \int_0^{\infty} \lambda e^{-\lambda x} (\lambda x)^s dx \\ &= \frac{\Gamma(s+1)}{\lambda \Gamma(s)} \\ &= \frac{s}{\lambda}. \end{aligned}$$

Maintenant pour la variance, on a :

$$\begin{aligned}
 E(X^2) &= \frac{1}{\Gamma(s)} \int_0^\infty \lambda x^2 e^{-\lambda x} (\lambda x)^{s-1} dx \\
 &= \frac{1}{\lambda^2 \Gamma(s)} \int_0^\infty \lambda e^{-\lambda x} (\lambda x)^{s+1} dx \\
 &= \frac{\Gamma(s+2)}{\lambda^2 \Gamma(s)} \\
 &= \frac{(s+1)\Gamma(s+1)}{\lambda^2 \Gamma(s)} \\
 &= \frac{s(s+1)}{\lambda^2} \\
 \text{Var}(X) &= E(X^2) - (E(X))^2 \\
 &= \frac{s}{\lambda^2}.
 \end{aligned}$$

Lorsque le paramètre s est un nombre entier, on peut associer dans la plupart du temps au temps d'attente avant la $s^{\text{ième}}$ occurrence d'un événement quelconque.

2.6.3 Loi normale

La loi normale est une loi de probabilité très populaire depuis des années et est très utilisée pour évaluer certaines probabilités. Elle a été étudiée par le mathématicien français de Moivre, en 1733. La loi normale était souvent utilisée pour approximer des probabilités de loi binomiale dont le paramètre n était très grand. On dit qu'une variable aléatoire X de paramètres μ et σ^2 est normale si la densité $f(x)$ est donnée par

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}, \forall x \in \mathbb{R}.$$

Pour trouver l'espérance et la variance pour une loi normale μ et σ^2 , il faut étudier la loi normale centrée réduite. La loi centrée réduite est une variable aléatoire normale de paramètres $\mu = 0, \sigma^2 = 1$, on la nomme Z . Alors la densité sera :

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \forall z \in \mathbb{R}.$$

Pour passer de la loi normale à la normale centrée réduite, il suffit de poser le changement de variable $Z = (X - \mu)/\sigma$. Donc, on a :

$$E(Z) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} x e^{-x^2/2} dx = -\frac{1}{\sqrt{2\pi}} e^{-x^2/2} \Big|_{-\infty}^{\infty} = 0.$$

En intégrant par partie $\frac{1}{\sqrt{2\pi}}xe^{-x^2/2}$, on peut montrer que $\text{Var}(Z) = 1$. Puisque $X = \mu + \sigma Z$, $E(X) = \mu$ et $\text{Var}(X) = \sigma^2$. La fonction de répartition n'est pas définie puisqu'il n'existe pas de primitive à la densité. Alors, il existe une table pour la loi normale centrée réduite qui permet d'exprimer une approximation de l'aire sous la courbe cumulée représentée par $\alpha(x)$. Pour approximer $\alpha(x)$, il suffit d'approximer :

$$\alpha(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}}e^{-y^2/2}dy.$$

Puisque la loi normale est symétrique en μ , on a que $\alpha(0) = \frac{1}{2}$ et $\alpha(-x) = 1 - \alpha(x)$. Alors, on a $P(Z \leq X) = \alpha(x)$. Si la loi normale n'est pas centrée réduite, il faut passer par le changement de variable $Z = (X - \mu)/\sigma$.

2.7 Définition d'un vecteur aléatoire continu

Un vecteur aléatoire continu $X = (X_1, X_2, \dots, X_n)$ est un vecteur de \mathbb{R}^n telle que pour tout $i \in [1, n]$, X_i est une variable aléatoire continue. Il existe, pour ces vecteurs, une fonction de répartition conjointe ou simultanée à une paire de variables. Pour débiter, prenons un vecteur aléatoire avec une paire de variable aléatoire X et Y .

Définition 2.7.1. Soit X et Y deux variables aléatoires continues. On nomme F la fonction de répartition conjointe de X, Y tel que pour tous $a, b \in (-\infty, \infty)$:

$$F(a, b) = P(X \leq a, Y \leq b)$$

On peut généraliser avec X_1, X_2, \dots, X_n variables aléatoires dans \mathbb{R}^n . Cette définition est directement liée à la définition de la fonction de répartition d'une seule variable aléatoire. Le but d'étudier plusieurs variables aléatoires est de tenter de caractériser le lien simultanément. On peut déduire, par la suite, les fonctions de répartitions de chacune des variables aléatoires.

Définition 2.7.2. Les fonctions de répartitions marginales de X et de Y sont déduites par la fonction de répartition jointe telle que :

$$F_X(a) = F(a, \infty) \text{ et } F_Y(b) = F(\infty, b).$$

On peut faire cette déduction puisque :

$$\begin{aligned}
 F_X(a) &= P(X \leq a) \\
 &= P(X \leq a, Y < \infty) \\
 &= P\left(\lim_{b \rightarrow \infty} (X \leq a, Y \leq b)\right) \\
 &= \lim_{b \rightarrow \infty} P(X \leq a, Y \leq b) \\
 &= \lim_{b \rightarrow \infty} F(a, b) = F(a, \infty).
 \end{aligned}$$

2.8 Densité d'un vecteur aléatoire continu

La densité d'un vecteur aléatoire continu est étroitement liée à la densité d'une variable aléatoire continue. Elle permet de trouver certaines probabilités. Par exemple, on peut s'interroger sur la valeur de $P(X \geq 1, Y \leq 1)$. Il suffit de trouver une définition dans le cas d'une paire de variables aléatoires continues.

Définition 2.8.1. *Dans le cas d'une paire de variables aléatoires X et Y , telles que les deux sont non-dénombrables, on définira la densité conjointe ou simultanée, la fonction f à deux variables telle que, pour tout sous ensemble A du plan de \mathbb{R}^2 :*

$$P\{(X, Y) \in A\} = \int \int_{(x,y) \in A} f(x, y) dx dy.$$

Tout comme pour la densité d'une variable aléatoire continue, pour s'assurer que $f(x, y)$ est bel et bien une densité conjointe, il faut :

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1.$$

La densité conjointe de X et Y peut se généraliser à plusieurs variables. Il suffit d'utiliser la fonction de répartition conjointe des n variables X_1, \dots, X_n . Pour n variables, on aura n intégrales à résoudre et il faudra que l'intégration de la densité conjointe $f(x_1, \dots, x_n)$ dans $\mathbb{R}^n = 1$ également.

Exemple 2.8.1. *Soit X et Y une paire de variables aléatoires continues telle que la densité conjointe est donnée par :*

$$f(x) = \begin{cases} ke^{-x}e^{-2y} & \text{si } x > 0, y > 0, \\ 0 & \text{sinon.} \end{cases}$$

Trouvons k tel que f est une densité. Utilisons le théorème de Fubini pour l'intégrale double, puisque $f(x, y) = f(x)g(y)$

$$\begin{aligned}
 1 &= \int_0^\infty \int_0^\infty k e^{-x} e^{-2y} dx dy &= k \int_0^\infty \int_0^\infty e^{-x} e^{-2y} dx dy \\
 & &= k \int_0^\infty e^{-x} dx \int_0^\infty e^{-2y} dy \\
 & &= k \left(-e^{-x} \Big|_0^\infty \right) \left(\frac{-e^{-2y}}{2} \Big|_0^\infty \right) \\
 & &= k (1) \left(\frac{1}{2} \right) = \frac{k}{2} \\
 &\Rightarrow k &= 2.
 \end{aligned}$$

Alors, la densité conjointe entre X et Y est :

$$f(x, y) = \begin{cases} 2e^{-x}e^{-2y} & \text{si } x > 0, y > 0, \\ 0 & \text{sinon.} \end{cases}$$

Par cette définition, il est possible de déterminer la probabilité

$$P(X \in A, Y \in B) = \int_A \int_B f(x, y) dx dy.$$

En utilisant la définition 2.8.1, on peut déduire :

$$\begin{aligned}
 F(a, b) &= P(X \leq a, Y \leq b) \\
 &= P(X \in [-\infty, a], Y \in [-\infty, b]) \\
 &= \int_{-\infty}^b \int_{-\infty}^a f(x, y) dx dy.
 \end{aligned}$$

La densité d'une variable aléatoire s'obtient à partir de la dérivée de la fonction de répartition. Pour une paire de variables aléatoires X et Y , on dérive pour obtenir :

$$f(a, b) = \frac{\partial^2}{\partial a \partial b} F(a, b).$$

2.9 Densité marginale d'un vecteur aléatoire continu

Tout comme la fonction de répartition marginale, il existe la densité marginale d'un vecteur aléatoire continu. Pour une paire de variables aléatoires (X, Y) , on peut déduire

deux densités marginales, soit une par rapport à X et une par rapport à Y . Le calcul pour la densité marginale par rapport à y est fait de la même façon par rapport à celle en x . On obtient comme densité marginale :

$$\begin{aligned}
 P(X \in A) &= P(X \in A, Y \in \mathbb{R}) \\
 &= \int_A \int_{-\infty}^{\infty} f(x, y) dy dx \\
 &= \int_A f_X(x) dx \\
 \Rightarrow f_X(x) &= \int_{-\infty}^{\infty} f(x, y) dy \\
 \Rightarrow f_Y(y) &= \int_{-\infty}^{\infty} f(x, y) dx.
 \end{aligned}$$

Exemple 2.9.1. Reprenons la fonction de l'exemple précédent et trouvons les densités marginales associées à la densité trouvée précédemment. On avait comme fonction de densité :

$$f(x, y) = \begin{cases} 2e^{-x}e^{-2y} & \text{si } x > 0, y > 0, \\ 0 & \text{sinon.} \end{cases}$$

Calculons, pour débiter, la densité marginale par rapport à X . On a donc :

$$\begin{aligned}
 f_X(x) &= \int_{-\infty}^{\infty} f(x, y) dy \\
 &= \int_{-\infty}^{\infty} 2e^{-x}e^{-2y} dy \\
 &= \int_{-\infty}^0 0 dy + \int_0^{\infty} 2e^{-x}e^{-2y} dy \\
 &= \int_0^{\infty} 2e^{-x}e^{-2y} dy \\
 &= 2e^{-x} \int_0^{\infty} e^{-2y} dy \\
 &= 2e^{-x} \left(\frac{-e^{-2y}}{2} \Big|_0^{\infty} \right) \\
 &= 2e^{-x} \left(0 - -\frac{1}{2} \right) \\
 &= e^{-x} \\
 f_Y(y) &= \int_0^{\infty} 2e^{-x}e^{-2y} dx \\
 &= 2e^{-2y}.
 \end{aligned}$$

2.10 Exemples de vecteurs aléatoires continus

Il existe une multitude de combinaisons de vecteurs aléatoires continus possibles. La loi multidimensionnelle la plus classique est la loi normale multidimensionnelle. Dans

cette section, on traitera le cas de la loi normale bidimensionnelle.

Tout comme dans le contexte univarié, on peut définir de façon semblable la loi normale bidimensionnelle par sa densité.

Définition 2.10.1. Soit X et Y , deux variables aléatoires normales telles que $X \sim \mathcal{N}(\mu_x, \sigma_x^2)$ et $Y \sim \mathcal{N}(\mu_y, \sigma_y^2)$ et ρ le coefficient de corrélation entre X et Y alors, la densité de probabilité conjointe est définie telle que :

$$f(x, y) = \frac{\exp(A)}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}.$$

$$\text{Tel que } A = \frac{-1}{2} \left[\left(\frac{x - \mu_x}{\sigma_x} \right)^2 + \left(\frac{y - \mu_y}{\sigma_y} \right)^2 - 2\rho \frac{(x - \mu_x)(y - \mu_y)}{(1 - \rho^2)\sigma_x\sigma_y} \right].$$

Lorsque les lois normales sont des lois normales centrées réduites $X \sim \mathcal{N}(0, 1)$, $Y \sim \mathcal{N}(0, 1)$, alors la densité jointe sera alors :

$$f(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp \left\{ \frac{-1}{2} \left(x^2 + y^2 - \frac{2\rho xy}{1-\rho^2} \right) \right\}.$$

Chapitre 3

Rappel sur les copules

Dans ce chapitre, nous étudierons certaines notions élémentaires des copules. Les copules modélisent une structure de dépendance entre les variables X et Y à partir des fonctions de distributions vues au chapitre précédent. Le théorème de Sklar qui est fondamental pour l'étude des copules sera étudié. Par la suite, à l'aide d'une copule, nous construirons une loi multidimensionnelle et à partir de celle-ci on pourra extraire une copule. Pour finir, nous étudierons deux mesures de concordances soit le tau de Kendall et le Rho de Spearman. Nous étudierons également quelques copules classiques telles que les copule archimédiennes, normale et de Student. La plupart des résultats présentés ici sont repris du livre de Ross (5).

3.1 Théorème de Sklar

On a vu précédemment que deux variables aléatoires X et Y tel que :

$$F(X) = P(X \leq x) \text{ et } G(y) = P(Y \leq y).$$

possèdent une loi de distribution $H(x, y)$ tel que :

$$H(x, y) = P(X \leq x, Y \leq y).$$

Le théorème de Sklar, établie en 1959, par Abe Sklar (9), isolera la dépendance entre X et Y . C'est le théorème fondamental pour la définition d'une copule.

Théorème 3.1.1. *Soit H la loi de distribution jointe décrite par les marges F et G telle que définie précédemment. Alors, il existe une copule C tel que $\forall (x, y) \in \mathbb{R}^2$,*

$$H(x, y) = C(F(x), G(y)).$$

Il faut noter que la copule est une fonction de dépendance. Elle ne dépend pas des lois marginales. Montrons qu'avec $U = F(X)$ et $V = G(Y)$, la copule C correspond au vecteur aléatoire (U, V) tel que $F(X) \sim \mathcal{U}_{[0,1]}$ et $G(Y) \sim \mathcal{U}_{[0,1]}$.

Démonstration.

$$\begin{aligned} H(x, y) &= P(F(X) \leq u, G(Y) \leq v) \\ &= P(X \leq F^{-1}(u), Y \leq G^{-1}(v)) \\ &= H(F^{-1}, G^{-1}) \\ &= C(F \circ F^{-1}(u), G \circ G^{-1}(v)) \\ &= C(u, v). \end{aligned}$$

□

3.2 Définition des copules

Exprimons une définition formelle de la copule de \mathbb{R}^2 à partir du théorème de Sklar énoncé précédemment.

Définition 3.2.1. *Une fonction $C : [0, 1] \times [0, 1] \rightarrow [0, 1]$ est une copule si et seulement s'il existe deux variables aléatoires*

$$U \sim \mathcal{U}_{[0,1]}, V \sim \mathcal{U}_{[0,1]}.$$

tel que

$$C(u, v) = P(U \leq u, V \leq v).$$

On peut aussi généraliser cette définition dans \mathbb{R}^n . Par le théorème de Sklar, on peut déduire quelques propriétés de base des copules telles que :

$$C(0, v) = C(u, 0) = 0,$$

$$C(1, u) = C(u, 1) = u.$$

$$C(u_2, v_2) - C(u_1, v_2) - C(u_2, v_1) + C(u_1, v_1) \geq 0, \quad \forall u_1 \leq u_2, v_1 \leq v_2.$$

3.3 Construction d'une loi multidimensionnelle à travers une copule

Une copule couple les lois marginales. Comme mentionné précédemment, la copule ne dépend pas des lois marginales. Le théorème de Sklar permet de construire la loi jointe H , avec les lois marginales de X et Y , puis d'une copule appropriée indépendante du choix des marges. Construisons deux distributions bivariées avec la même copule mais avec des lois marginales différentes.

Exemple 3.3.1. Soit la copule suivante :

$$C(u, v) = \frac{uv}{u + v - uv}.$$

Posons les marges $F(x) = G(x) = \sqrt{x}, \forall x \in [0, 1]$, on obtient directement la loi jointe

$$\begin{aligned} H(x, y) &= \frac{\sqrt{x}\sqrt{y}}{\sqrt{x} + \sqrt{y} - \sqrt{x}\sqrt{y}} \\ &= \frac{\sqrt{xy}}{\sqrt{x} + \sqrt{y} - \sqrt{xy}}. \end{aligned}$$

Par contre, avec les marges $F(x) = G(x) = 1 - \frac{x}{2}, \forall x \in [0, 2]$, la loi jointe obtenue est :

$$H(x, y) = \frac{(2-x)(2-y)}{2(2-xy)}.$$

3.4 Construction d'une copule à travers une distribution multidimensionnelle

Le théorème de Sklar permet aussi d'extraire une copule, à partir d'une loi jointe H . Puisque $u = F(x)$ et $v = G(y)$, on peut inverser les fonctions pour obtenir :

$$x = F^{-1}(u) \text{ et } y = G^{-1}(v).$$

En utilisant le théorème de Sklar, soit $H(x, y) = C(F(x), G(y))$, on obtient :

$$C(u, v) = H(F^{-1}(u), G^{-1}(v)).$$

Donc, à partir de la loi de $H(x, y)$, on peut déduire la copule associée.

Exemple 3.4.1. Soit $H(x, y)$ la loi jointe définie telle que :

$$H(x, y) = \sqrt{\frac{xy}{x + y - xy}}, \quad (x, y) \in [0, 1]^2.$$

Trouvons la copule associée à H . Pour ce faire, il faut déduire les lois marginales $F(x)$ et $G(y)$. On a que $H(x, 1) = F(x)$, car $H(x, 1) = P(F(X) \leq u, G(Y) \leq 1) = F(x)$. Par la suite, on aura $G(y) = H(1, y)$ pour la même raison que précédemment. Donc, on trouve comme lois marginales :

$$H(x, 1) = \sqrt{\frac{x}{x + 1 - x}} = \sqrt{x} = F(x),$$

$$H(1, y) = \sqrt{\frac{y}{1 + y - y}} = \sqrt{y} = G(y).$$

Maintenant, on peut inverser les deux lois marginales pour obtenir :

$$F(x) = \sqrt{x} \Rightarrow u = \sqrt{x} \Rightarrow u^2 = x = F^{-1}(u),$$

$$G(y) = \sqrt{y} \Rightarrow v = \sqrt{y} \Rightarrow v^2 = y = F^{-1}(v).$$

Maintenant, il suffit d'appliquer le théorème de Sklar. On obtient :

$$\begin{aligned} C(u, v) &= H(F^{-1}(u), G^{-1}(v)) \\ &= H(u^2, v^2) \\ &= \sqrt{\frac{u^2 v^2}{u^2 + v^2 - u^2 v^2}} \\ &= (u^{-2} + v^{-2} - 1)^{-1/2}. \end{aligned}$$

3.5 Copules de survie

Le théorème de Sklar s'utilise également dans le cas des fonctions de survie. Cela permet d'obtenir une expression de la survie d'un modèle quelconque. Il faut adapter le théorème pour obtenir une copule de survie. Les fonctions de survie sont donc :

$$\begin{aligned} H^*(x, y) &= P(X > x, Y > y) \\ F^*(x) &= P(X > x) \\ G^*(y) &= P(Y > y). \end{aligned}$$

Par le théorème de Sklar, il existe une copule $C^* : [0, 1] \times [0, 1] \rightarrow [0, 1]$ telle que pour tout $(x, y) \in \mathbb{R}^2$,

$$H^*(x, y) = C^*(F^*(x), G^*(y)).$$

Démonstration. Puisque $H^*(x, y) = P(X > x, Y > y)$ et en ajoutant un terme fantôme $-1 + 1$, on a que :

$$\begin{aligned} H^*(x, y) &= 1 - F(x) - G(y) + H(x, y) \\ &= 1 - 1 + 1 - F(x) - G(y) + H(x, y) \\ &= (1 - F(x)) + (1 - G(y)) - 1 + C(F(x), G(y)) \\ &= (F^*(x)) + (G^*(y)) - 1 + C(1 - (F^*(x)), 1 - G^*(y)). \end{aligned}$$

Il suffit de poser $C^*(u, v) = u + v - 1 + C(1 - u, 1 - v)$ pour obtenir le résultat désiré.

$$H^*(x, y) = C^*(F^*(x), G^*(y)).$$

C^* se nomme copule de survie. □

3.6 Quelques copules classiques

3.6.1 Copule Indépendante

Soit les variables aléatoires X et Y indépendantes, donc on a :

$$H(x, y) = C(F(x), G(y)) = F(x)G(y).$$

En posant $u = F(x)$ et $v = G(y)$, la copule précédente sera alors :

$$C(u, v) = uv.$$

Définition 3.6.1. Pour tout $u, v \in [0, 1]$, la copule d'indépendance est définie telle que :

$$\Pi(u, v) = uv$$

3.6.2 Bornes de Fréchet-Hoeffding

Théorème 3.6.1. *Soit une copule C quelconque, telle que $\forall u, v \in [0, 1]$:*

$$\max(u + v - 1, 0) \leq C(u, v) \leq \min(u, v).$$

Démonstration. Montrons que $C(u, v) \leq \min(u, v)$.

Soit un point $(u, v) \in [0, 1] \times [0, 1]$. On a alors par définition :

$$C(u, v) \leq C(u, 1) = u \text{ et } C(u, v) \leq C(1, v) = v.$$

Puisque $C(u, v) \leq u$ et $C(u, v) \leq v$, on a alors $C(u, v) \leq \min(u, v)$ et puisque $\max(u + v - 1, 0) \leq C(u, v) \leq \min(u, v)$ cela complète la démonstration. \square

Alors on associera $\max(u + v - 1, 0) = W(u, v)$ à la borne inférieure de Fréchet qui décrit la parfaite dépendance négative et $\min(u, v) = M(u, v)$ et la borne supérieure qui correspond à la parfaite dépendance positive.

3.6.3 Copule Archimédienne

Les copules archimédiennes ont été définies par Genest et Mackay (1) en 1986. Elles englobent plusieurs copules connues. Cette classe possède diverses applications distinctes. L'utilisation de ces copules est par sa simplicité à les construire. Une copule archimédienne de générateur ϕ est définie telle que :

Définition 3.6.2. *Soit ϕ une fonction décroissante défini de $[0, 1] \times [0, 1] \rightarrow [0, \infty[$ tel que $\phi(1) = 0$, alors une copule archimédienne s'exprime sous la forme :*

$$C_\phi(u, v) = \phi^{-1}\{\phi(u) + \phi(v)\}, \quad (u, v) \in [0, 1]^2.$$

ϕ^{-1} représente l'inverse de ϕ .

Montrons que ϕ est un unique générateur d'une copule archimédienne.

Démonstration. En effet, si ϕ génère une copule archimédienne C quelconque, alors $\forall k > 0$, on aurait $\Phi(t) = k\phi(t)$ aussi un générateur de C . Puisque $\Phi^{-1}(u) = \phi^{-1}\left(\frac{u}{k}\right)$, on a alors :

$$\begin{aligned} C_\Phi(u, v) &= \Phi^{-1}(\Phi(u) + \Phi(v)) \\ &= \Phi^{-1}(k(\phi(u) + \phi(v))) \\ &= \phi^{-1}(\phi(u) + \phi(v)) \\ &= C_\phi(u, v). \end{aligned}$$

Alors, on a que ϕ est unique. □

Les copules archimédiennes regroupent plusieurs grandes copules classiques. Maintenant par la définition précédente, on ne peut déterminer le générateur directement. Alors, le prochain théorème, démontré dans le livre de Ross (8), permettra de déterminer le générateur.

Théorème 3.6.2. *Soit C une copule archimédienne générée par ϕ , il est possible de déterminer le générateur en résolvant :*

$$\frac{\frac{\partial}{\partial u} C(u, v)}{\frac{\partial}{\partial v} C(u, v)} = \frac{\phi'(u)}{\phi'(v)}.$$

Utilisons cette technique pour déterminer le générateur d'une copule.

Exemple 3.6.1. *Montrons que le générateur de la copule ci-après s'exprime sous la forme suivante :*

$$C(u, v) = \frac{uv}{u + v - uv} \text{ avec } \phi(t) = \frac{1}{t} - 1.$$

Trouvons $\phi(t)$. Alors, on a :

$$\frac{\phi'(u)}{\phi'(v)} = \frac{\frac{\partial}{\partial u} C(u, v)}{\frac{\partial}{\partial v} C(u, v)} = \frac{v^2}{u^2}.$$

Puisque $\frac{\phi'(u)}{\phi'(v)} = \frac{v^2}{u^2}$, on constate que $u^2\phi'(u)$ est constant. Si $\phi'(u) < 0, \exists k_1 > 0$ tel que :

$$\phi'(u) = -\frac{k_1}{u^2}.$$

Intégrons $\phi'(u)$ par rapport à u .

$$\begin{aligned} \int \frac{-k_1}{u} du &= -k_1 \int \frac{du}{u^2} \\ &= -k_1 \int u^{-2} du \\ &= -k_1 \frac{u^{-1}}{-1} + C \\ &= \frac{k_1}{u} + k_2. \end{aligned}$$

Par la définition du générateur, il faut que $\phi(1) = 0$, alors on a,

$$\begin{aligned} \phi(u) &= \frac{k_1}{u} + k_2 \\ \Rightarrow \phi(1) &= \frac{k_1}{1} + k_2 \\ \Rightarrow 0 &= k_1 + k_2 \\ \Rightarrow k_1 &= -k_2 = k > 0 \\ \Rightarrow \phi(u) &= \frac{k}{u} + k \\ &= k \left(\frac{1}{u} - 1 \right), k > 0. \end{aligned}$$

Alors le générateur de la copule exprimée initialement est donc :

$$\phi(t) = \frac{1}{t} - 1.$$

3.6.4 Simulation de quelques copules classiques

La simulation des copules est très utilisée en pratique. La simulation de copule permet d'exprimées plusieurs modèles différents. À l'aide du logiciel R Studio, cette section décrira comment simuler un jeu de données à partir d'une copule quelconque. Pour simuler une copule, on doit exprimer une méthode fiable pour s'assurer que toutes les copules soient simulées. Le théorème qui suit exprime un algorithme qui génère des données à partir d'une copule C .

Théorème 3.6.3. Soit C une copule quelconque et (U, V) un vecteur aléatoire uniforme avec C . Alors pour tous $(u, v) \in [0, 1]^2$, on a :

$$L_v(u) = P(U \leq u | V = v) = \frac{\partial}{\partial v} C(u, v).$$

Soit G_1 et G_2 deux variables aléatoires indépendantes telles que $G_1 \sim \mathcal{U}_{[0,1]}$ et $G_2 \sim \mathcal{U}_{[0,1]}$. Posons $X = L_{G_2}^{-1}(G_1)$ et $Y = G_2$. Alors, la copule de la paire aléatoire est C .

Démonstration. Puisque G_1 et G_2 sont indépendantes, $G_1 \sim \mathcal{U}_{[0,1]}$ et $C(u, 0) = 0$, alors on a :

$$\begin{aligned} P(X \leq u, Y \leq v) &= P(L_{G_2}^{-1}(G_1) \leq u, G_2 \leq v) \\ &= \int_0^1 P(L_{G_2}^{-1}(G_1) \leq u, G_2 \leq v | G_2 = s) ds \\ &= \int_0^v P(L_s^{-1}(G_1) \leq u) ds \\ &= \int_0^v L_s(u) ds \\ &= \int_0^v \frac{\partial}{\partial s} C(u, s) ds \\ &= C(u, v). \end{aligned}$$

□

Alors, la copule de la paire aléatoire (X, Y) tel que $X = L_{G_2}^{-1}(G_1)$ et $Y = G_2$ est C . À partir du dernier théorème, on établira un algorithme permettant de simuler les copules qui suit :

Définition 3.6.3. *L'algorithme permettant de générer les copules est :*

1. Générer les variables indépendantes u, t de loi uniforme $[0, 1]$
2. Posons $v = L_u^{-1}(t)$.
3. Le couple désiré sera (v, w)

Pour utiliser l'algorithme, il faut trouver le terme $L_v^{-1}(u)$. Dans le cas des copules archimédiennes, on suppose C une copule archimédienne générée par ϕ ,

$$C(u, v) = \phi^{-1}(\phi(u) + \phi(v)).$$

Alors, trouvons $L_v^{-1}(u)$. Pour ce faire, il suffit de débiter en trouvant $L_v(u) = \frac{\partial}{\partial v} C(u, v)$.

Donc, on a, à l'aide d'un logiciel de calcul :

$$\begin{aligned}\frac{\partial}{\partial v} C(u, v) &= \frac{\phi'(v)}{\phi'(\phi^{-1}(\phi(u) + \phi(v)))} \\ \Rightarrow L_v^{-1}(u) &= \phi^{-1} \left(\phi \left((\phi')^{-1} \left(\frac{\phi'(v)}{t} \right) \right) - \phi(v) \right).\end{aligned}$$

Donc, on peut maintenant écrire un algorithme permettant de simuler les copules archimédiennes :

- Générer les variables indépendantes u, t de loi uniforme $[0, 1]$
- Posons $v = L_u^{-1}(t) = \phi^{-1}(\phi((\phi')^{-1}(\frac{\phi'(v)}{t})) - \phi(v))$
- Alors, la paire désirée est donc (u, v) .

Passons maintenant à la simulation de copules connues. La copule de Clayton et la copule de Frank ayant déjà été mentionnées précédemment seront utilisées à titre d'exemples.

Exemple 3.6.2. *Simulons la copule de Clayton avec diverses valeurs de θ . Premièrement, trouvons $L_u^{-1}(t)$. Précédemment, on a montré que $\phi(t) = \frac{t^{-\theta} - 1}{\theta}$. Trouvons l'inverse de $\phi(t)$. On a :*

$$\begin{aligned}\phi(t) &= \frac{t^{-\theta} - 1}{\theta} \\ \Rightarrow s &= \frac{t^{-\theta} - 1}{\theta} \\ \Rightarrow \theta s &= t^{-\theta} - 1 \\ \Rightarrow \theta s + 1 &= t^{-\theta} \\ \Rightarrow t &= (\theta s + 1)^{-1/\theta} \\ \Rightarrow \phi^{-1}(s) &= (\theta s + 1)^{-1/\theta}.\end{aligned}$$

Par la suite, trouvons $\phi'(t)$ et $(\phi')^{-1}(s)$.

$$\begin{aligned}\frac{d}{dt} \left(\frac{t^{-\theta} - 1}{\theta} \right) &= \frac{1}{\theta} \frac{d}{dt} (t^{-\theta}) - \frac{d}{dt} (1) \\ &= \frac{1}{\theta} \frac{d}{dt} (t^{-\theta}) \\ &= -\frac{\theta}{\theta} t^{-\theta-1} \\ &= -t^{-\theta-1} \\ \Rightarrow (\phi')^{-1}(s) &= (-s)^{-1/\theta+1}. \\ \Rightarrow L_u^{-1}(t) &= \left(u^{-\theta} t^{-\theta/(\theta+1)} - u^{-\theta} + 1 \right)^{-1/\theta}.\end{aligned}$$

Donc l'algorithme permettant de générer des données à partir de la copule de Clayton est :

- Générer les variables indépendantes u, t de loi uniforme $[0, 1]$
- Posons $v = L_u^{-1}(t) = \left(u^{-\theta} t^{-\theta/(\theta+1)} - u^{-\theta} + 1\right)^{-1/\theta}$
- Alors, la paire désirée est donc (u, v) .

Les deux figures ci-contre montre le cas d'une dépendance négative ($\theta < 0$) et le cas d'une dépendance positive ($\theta > 0$).

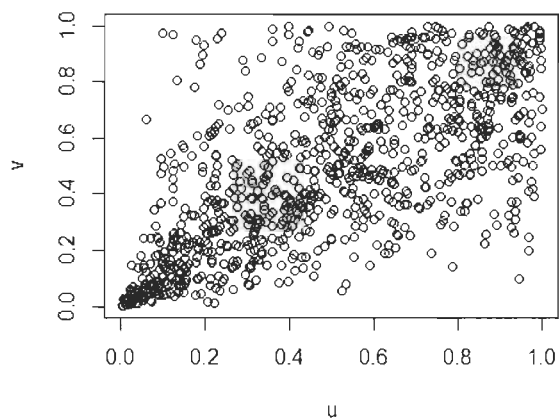


FIGURE 3.1 – Simulation de la copule de Clayton avec $\theta = 2$.

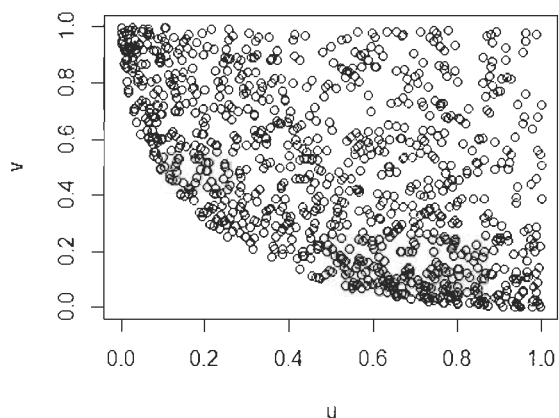


FIGURE 3.2 – Simulation de la copule de Clayton avec $\theta = -2$.

Exemple 3.6.3. *Tentons de développer l'algorithme de simulation pour la copule de Frank à l'aide du théorème pour les copules archimédiennes. Puisque la copule de Frank est une copule archimédienne, on a comme générateur de la copule de Frank :*

$$\phi(t) = -\ln \left(\frac{1 - e^{-t}}{1 - e^{-\theta t}} \right).$$

À l'aide d'un logiciel de calcul, on obtient :

$$L_u^{-1}(t) = \frac{1}{\theta} \ln \left(\frac{e^{-\theta u} + t(1 - e^{-\theta u})}{e^{\theta u} + t(e^{-\theta} - e^{-\theta u})} \right).$$

Par la suite, on peut exprimer l'algorithme de génération de données à partir de la copule de Frank qui suit.

— Générer les variables indépendantes u, t de loi uniforme $[0, 1]$

— Posons $v = \frac{1}{\theta} \ln \left(\frac{e^{-\theta u} + t(1 - e^{-\theta u})}{e^{\theta u} + t(e^{-\theta} - e^{-\theta u})} \right)$

— Alors, la paire désirée est donc (u, v) .

Les deux figures ci-contre montrent le cas d'une dépendance négative ($\theta < 0$) et le cas d'une dépendance positive ($\theta > 0$).

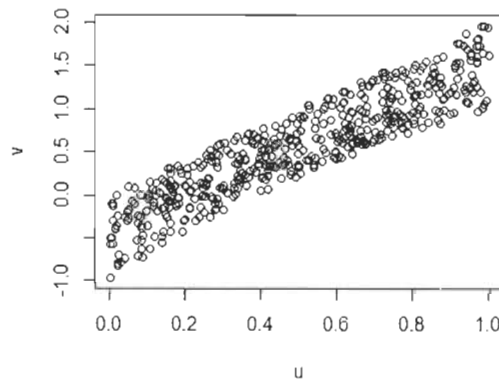


FIGURE 3.3 – Simulation de la copule de Frank avec $\theta = 0.5$

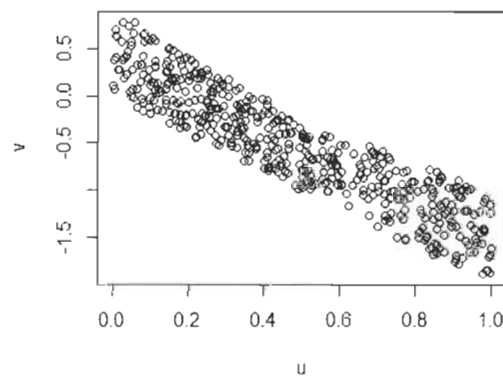


FIGURE 3.4 – Simulation de la copule de Frank avec $\theta = -0.5$

3.7 Mesures de concordance : Tau de Kendall et rho de Spearman

Cette section traitera de l'étude de la dépendance à l'aide des copules. Deux mesures très utilisées seront explorées, le tau de Kendall et le rho de Spearman. Par ailleurs, il faut débiter par définir ce qu'est la concordance et la discordance.

Définition 3.7.1. Soit X et Y une paire de variables aléatoires (X, Y) . Soit les couples (x_i, y_i) et (x_j, y_j) . On dit que les couples (x_i, y_i) et (x_j, y_j) sont concordants si $x_i < x_j$

et $y_i < y_j$ ou si $x_i > x_j$ et $y_i > y_j$. Par opposition, on dit que (x_i, y_i) et (x_j, y_j) sont *disconcordants* si $x_i < x_j$ et $y_i > y_j$ ou $x_i > x_j$ et $y_i < y_j$.

Trouvons une formulation directe de la définition précédente.

$$\begin{aligned} & x_i < x_j \text{ et } y_i < y_j \text{ ou } x_i > x_j \text{ et } y_i > y_j. \\ \Rightarrow & x_i - x_j < 0 \text{ et } y_i - y_j < 0 \text{ ou } x_i - x_j > 0 \text{ et } y_i - y_j > 0. \\ \Rightarrow & (x_i - x_j)(y_i - y_j) > 0 \text{ ou } (x_i - x_j)(y_i - y_j) > 0. \\ \Rightarrow & (x_i - x_j)(y_i - y_j) > 0. \end{aligned}$$

De la même façon, on a pour la disconcordance :

$$\begin{aligned} & x_i < x_j \text{ et } y_i > y_j \text{ ou } x_i > x_j \text{ et } y_i < y_j. \\ \Rightarrow & x_i - x_j < 0 \text{ et } y_i - y_j > 0 \text{ ou } x_i - x_j > 0 \text{ et } y_i - y_j < 0. \\ \Rightarrow & (x_i - x_j)(y_i - y_j) < 0 \text{ ou } (x_i - x_j)(y_i - y_j) < 0. \\ \Rightarrow & (x_i - x_j)(y_i - y_j) < 0. \end{aligned}$$

À l'aide de la concordance entre deux couples de variables aléatoires, on peut calculer les concordances pour n couples. Alors, explorons les deux mesures de concordances les plus utilisées pour l'étude des copules : Le tau de Kendall et le rho de Spearman.

3.7.1 Tau de Kendall

Le tau de Kendall se définit pour les n observations générées par les variables aléatoires. On cherche à définir le tau de Kendall par la proportion de la différence entre le nombre de concordances et des disconcordances avec la somme des concordances et des disconcordances. Alors, on définit le tau de Kendall tel que :

Définition 3.7.2. Soit n observations générant n couples de variables aléatoires (x_i, y_i) possédant c concordances et d disconcordances, alors le Tau de Kendall est exprimé tel que :

$$t = \frac{c - d}{c + d}.$$

Puisque $c + d = \binom{2}{n}$, causé par le nombre de couplage des n observations deux à deux de 2 parmi n façons, alors on a :

$$t = \frac{c - d}{\binom{2}{n}} = \frac{c}{\binom{2}{n}} - \frac{d}{\binom{2}{n}}.$$

Donc, on dit que le tau de Kendall est calculée à partir de la probabilité d'obtenir une concordance moins la probabilité d'obtenir une discordance. On peut conclure que le tau de Kendall $-1 \leq t \leq 1$. Alors, on a défini le tau de Kendall pour le cas où les variables aléatoires sont discrètes. Par la suite, définissons le tau de Kendall dans le cas continu.

Définition 3.7.3. Soit (X, Y) un vecteur aléatoire de fonction de répartition H . Soient (X_1, Y_1) et (X_2, Y_2) , deux vecteurs aléatoires identiquement distribués de loi jointe H . Le tau de Kendall associé à H est défini par :

$$\tau_H = P((x_1 - x_2)(y_1 - y_2) > 0) - P((x_1 - x_2)(y_1 - y_2) < 0).$$

Notons que $-1 \leq \tau \leq 1$. $\tau_H = 0$ si X et Y sont indépendantes.

3.7.2 Rho de Spearman

Tout comme le tau de Kendall, le rho de Spearman est une mesure qui étudie la concordance ainsi que la discordance. Par contre, pour établir cette mesure, on étudiera trois vecteurs aléatoires. Définissons plus concrètement le rho de Spearman.

Définition 3.7.4. Soit $(X_1, Y_1), (X_2, Y_2), (X_3, Y_3)$, trois vecteurs aléatoires indépendantes de loi jointe H et de copule C . Le rho de Spearman ρ est proportionnel à la différence entre la concordance et la discordance entre la paire de vecteur avec les mêmes lois marginales, donc un distribué selon H c'est-à-dire que (X_1, Y_1) et (X_2, Y_3) et les autres composantes sont indépendantes :

$$\rho = 3(P((X_1 - X_2)(Y_1 - Y_3) > 0) - P((X_1 - X_2)(Y_1 - Y_3) < 0)).$$

Le facteur de 3 permet de normaliser les bornes, car ρ serait situé entre $[-1/3, 1/3]$.

3.8 Expression du Tau de Kendall et Rho de Spearman de la copule sous-jacente

Le tau de Kendall associé à (X, Y) de copule C s'exprime en fonction de cette dernière comme suit :

$$\begin{aligned}\tau &= 4 \int_0^1 \int_0^1 C(u, v) \frac{\partial^2 C}{\partial u \partial v} du dv - 1 \\ &= 4E(C(u, v)) - 1.\end{aligned}$$

Dans le cas d'une copule archimédienne, le tau de Kendall s'exprime en fonction du générateur ϕ .

Puisque X_2 et Y_3 sont indépendantes, la loi jointe est donc $F(x)G(y)$ et la copule est alors indépendante Π . Alors on peut déterminer le rho de Spearman de cette façon :

$$\rho = 12 \int_0^1 \int_0^1 C(u, v) du dv - 3 = 3Q(C, \Pi).$$

Théorème 3.8.1. *Soit X et Y deux variables aléatoires sous une copule archimédienne C générée par ϕ , le tau de Kendall est alors :*

$$t_C = 1 + \int_0^1 \frac{\phi(t)}{\phi'(t)} dt.$$

Démonstration. Soit u, v des lois uniformes $(0, 1)$ de loi jointe C et K_c la loi de distribution de $C(u, v)$. Alors, précédemment par la définition de $K_c(t)$, on a :

$$\begin{aligned}\tau_c &= 4 \int_0^1 \int_0^1 C(u, v) du dv - 1 \\ &= 4 \int_0^1 t K_c(t) dt - 1.\end{aligned}$$

En intégrant par partie, on obtient donc :

$$\begin{aligned}\tau_c &= 3 - 4 \int_0^1 K_c(t) dt \\ &= 3 - 4 \int_0^1 t - \frac{\phi(t)}{\phi'(t)} dt \\ &= 3 - 4 \int_0^1 t dt + 4 \int_0^1 \frac{\phi(t)}{\phi'(t)} dt \\ &= 3 - \frac{4t^2}{2} \Big|_0^1 + 4 \int_0^1 \frac{\phi(t)}{\phi'(t)} dt \\ &= 1 + 4 \int_0^1 \frac{\phi(t)}{\phi'(t)} dt.\end{aligned}$$

□

Exemple 3.8.1. Reprenons l'exemple de la copule de Clayton comme dans les exemples précédents. On doit se rappeler que la copule de Clayton est générée par $\phi_\theta(t) = \frac{t^{-\theta} - 1}{\theta}$. Puisque la copule de Clayton est une copule archimédienne, on peut appliquer le théorème précédent. Premièrement, on a :

$$\begin{aligned}\phi(t) &= \frac{t^{-\theta} - 1}{\theta}, \\ \Rightarrow \phi'(t) &= -t^{-\theta-1}.\end{aligned}$$

Calculons $\frac{\phi(t)}{\phi'(t)}$. On a :

$$\begin{aligned}\frac{\phi(t)}{\phi'(t)} &= \frac{\frac{t^{-\theta} - 1}{\theta}}{-t^{-\theta-1}} \\ &= \frac{t^{-\theta} - 1}{\theta t^{-\theta-1}} \\ &= \frac{-(t^{-\theta} - 1)t^{\theta+1}}{\theta} \\ &= \frac{-(t^{-\theta} - 1)t^{\theta+1}}{\theta} \\ &= \frac{-(t^{-\theta+\theta+1} - t^{\theta+1})}{\theta} \\ &= \frac{t^{\theta+1} - t}{\theta}.\end{aligned}$$

Donc, calculons maintenant le tau de Kendall associé.

$$\begin{aligned}\tau_\theta &= 1 + 4 \int_0^1 \frac{t^{\theta+1} - t}{\theta} \\ &= 1 + \frac{4}{\theta} \int_0^1 t^{\theta+1} - \frac{4}{\theta} \int_0^1 t dt \\ &= 1 + \frac{4}{\theta} \frac{t^{\theta+2}}{(\theta+2)} \Big|_0^1 - \frac{4}{\theta} \frac{t^2}{2} \Big|_0^1 \\ &= 1 + \frac{4}{\theta(\theta+2)} - \frac{2}{\theta} \\ &= \frac{\theta(\theta+2) + 4 - 2(\theta+2)}{\theta(\theta+2)} \\ &= \frac{\theta^2 + 2\theta + 4 - 2\theta - 4}{\theta(\theta+2)} \\ &= \frac{\theta^2}{\theta(\theta+2)} \\ &= \frac{\theta}{\theta+2}.\end{aligned}$$

Alors le tau de Kendall associé à la copule de Frank est alors $\frac{\theta}{\theta + 2}$.

Passons maintenant à un exemple du calcul du rho de Spearman d'une copule.

Exemple 3.8.2. Soit C_θ une copule de Farlie-Gumbel-Morgenstern définie pour $\theta \in [-1, 1]$ tel que :

$$C_\theta(u, v) = uv + \theta uv(1 - v)(1 - u).$$

Trouvons ρ_c .

$$\rho_c = 12 \int_0^1 \int_0^1 C(u, v) du dv - 3.$$

Commençons par résoudre l'intégrale. Alors, on a :

$$\begin{aligned} I &= \int_0^1 \int_0^1 uv + \theta uv(1 - v)(1 - u) du dv \\ &= \int_0^1 \int_0^1 uv du dv + \theta \int_0^1 \int_0^1 uv(1 - v)(1 - u) du dv \\ &= \int_0^1 v \int_0^1 u du dv + \theta \int_0^1 \int_0^1 uv(1 - v)(1 - u) du dv \\ &= \int_0^1 \frac{v}{2} + \theta \int_0^1 v(1 - v) \int_0^1 u(1 - u) du dv \\ &= \frac{1}{4} + \theta \int_0^1 v(1 - v) \int_0^1 (u - u^2) du dv \\ &= \frac{1}{4} + \theta \int_0^1 v(1 - v) \left(\frac{u^2}{2} - \frac{u^3}{3} \right) \Big|_0^1 dv \\ &= \frac{1}{4} + \theta \int_0^1 v(1 - v) \left(\frac{1}{6} \right) \\ &= \frac{1}{4} + \left(\frac{\theta}{6} \right) \int_0^1 v(1 - v) \\ &= \frac{1}{4} + \frac{\theta}{36}. \end{aligned}$$

Chapitre 4

Nouvelle loi exponentielle bidimensionnelle de type Raftery

Raftery et al (6) (1984) a proposé un modèle exponentiel bidimensionnel en utilisant un principe de choc commun particulier qui permet de créer la dépendance entre des composantes exponentielles indépendantes. Ce modèle présente un désavantage dans le sens qu'il permet de modéliser seulement les paires de variables aléatoires exponentielles identiquement distribuées. Ceci vient du fait que la dépendance est créée par une seule variable, appelée choc commun. Pour remédier à ce problème, nous proposons dans ce chapitre une loi exponentielle bidimensionnelle capable de modéliser des couples de lois exponentielles qui ne sont pas nécessairement identiquement distribuées. La procédure proposée se base sur le principe de chocs comonotones introduit récemment par Genest, Mesfioui et Schulz (2) (2018) dans le contexte de lois de Poisson.

4.1 Modèle exponentiel bidimensionnel de Raftery

Raftery et al (6) (1984) a suggéré une loi exponentielle bidimensionnelle définie comme suit.

Définition 4.1.1. Soient Z_1, Z_2 et Z_3 , des variables aléatoires indépendantes identiquement distribuées de loi exponentielle de paramètre λ , et J qui suit une loi de Bernoulli de paramètre θ qui est indépendante de Z_1, Z_2 et Z_3 . Le couple de variables aléatoires

défini par

$$\begin{aligned} X &= (1 - \theta)Z_1 + JZ_3, \\ Y &= (1 - \theta)Z_2 + JZ_3, \end{aligned}$$

suit une loi exponentielle bidimensionnelle de Raftery.

Raftery et al (7)(1984) a montré que les variables aléatoires marginales X et Y suivent des lois exponentielles du même paramètre λ . Ce qui montre que ce modèle peut être limité en pratique.

4.2 Nouveau modèle exponentiel bidimensionnel de type Raftery

Dans cette section, nous proposons un modèle exponentiel bidimensionnel permettant d'étendre le modèle Raftery introduit précédemment au cas général où les variables aléatoires marginales ne sont pas identiquement distribuées.

Définition 4.2.1. Soient Z_1, Z_2, Z , des variables aléatoires indépendantes telles que $Z_1 \sim \text{Exp}(\lambda_1)$, $Z_2 \sim \text{Exp}(\lambda_2)$ et $Z \sim \text{Exp}(1)$ et J , qui suit une loi de Bernoulli de paramètre θ qui est indépendante de Z_1, Z_2 et Z . La famille $\mathcal{E}(\theta, \lambda_1, \lambda_2)$ de variables aléatoires défini par

$$\begin{aligned} X &= (1 - \theta)Z_1 + J\frac{Z}{\lambda_1}, \\ Y &= (1 - \theta)Z_2 + J\frac{Z}{\lambda_2}, \end{aligned}$$

est appelée famille de lois exponentielles bidimensionnelles de type Raftery généralisée.

L'indice $\theta \in [0, 1]$ peut être vu comme paramètre de dépendance, car il n'affecte pas les lois marginales de X et Y comme le montre la proposition suivante.

Proposition 4.2.1. Les variables X et Y définies ci-dessus suivent respectivement des lois exponentielles de paramètres λ_1 et λ_2 . En d'autres termes, les fonction de survie de X et Y sont de la forme :

$$P(X > x) = \exp(-\lambda_1 x) \quad \text{et} \quad P(Y > x) = \exp(-\lambda_2 x) \quad \text{pour tout } x \geq 0.$$

Démonstration. Posons $Z_3 = Z/\lambda_1$. Il est clair que $Z_3 \sim \text{Exp}(\lambda_1)$. Puisque les variables aléatoires Z_1, Z_3 et J sont indépendantes, donc :

$$\begin{aligned}
 P(X > x) &= P[(1 - \theta)Z_1 + JZ_3] \\
 &= P[(1 - \theta)Z_1 + JZ_3 | J = 0]P(J = 0) \\
 &\quad + P[(1 - \theta)Z_1 + JZ_3 | J = 1]P(J = 1) \\
 &= (1 - \theta)P[(1 - \theta)Z_1 + JZ_3 | J = 0] \\
 &\quad + \theta P[(1 - \theta)Z_1 + JZ_3 | J = 1] \\
 &= (1 - \theta)P\left(Z_1 > \frac{x}{1 - \theta}\right) \\
 &\quad + \theta P((1 - \theta)Z_1 + Z_3 > x).
 \end{aligned}$$

Le premier terme de l'équation précédente est clairement exprimé par :

$$(1 - \theta)P\left(Z_1 > \frac{x}{1 - \theta}\right) = (1 - \theta) \exp\left(\frac{-\lambda_1 x}{1 - \theta}\right).$$

Il reste à calculer le second terme. En effet, nous avons

$$\begin{aligned}
 \theta P[(1 - \theta)Z_1 + Z_3 > x] &= \theta \int_0^\infty P[(1 - \theta)Z_1 + t > x | Z_3 = t] f_{Z_3}(t) dt \\
 &= \theta \int_0^\infty P[(1 - \theta)Z_1 + t > x] f_{Z_3}(t) dt \\
 &= \theta \int_0^\infty P[(1 - \theta)Z_1 + t > x] \lambda_1 e^{-\lambda_1 t} dt \\
 &= \theta \int_0^\infty P[(1 - \theta)Z_1 > x - t] \lambda_1 e^{-\lambda_1 t} dt \\
 &= \theta \int_0^x P\left(Z_1 > \frac{x - t}{1 - \theta}\right) \lambda_1 e^{-\lambda_1 t} dt + \theta \int_x^\infty \lambda_1 e^{-\lambda_1 t} dt \\
 &= \theta \int_0^x P\left(Z_1 > \frac{x - t}{1 - \theta}\right) \lambda_1 e^{-\lambda_1 t} dt + \theta e^{-\lambda_1 x}.
 \end{aligned}$$

Maintenant calculons à part l'intégrale impliquée dans la dernière équation.

$$\begin{aligned}
\theta \int_0^x P\left(Z_1 > \frac{x-t}{1-\theta}\right) \lambda_1 e^{-\lambda_1 t} dt &= \theta \int_0^x \exp\left(\frac{-\lambda_1(x-t)}{1-\theta}\right) \lambda_1 e^{-\lambda_1 t} dt \\
&= \lambda_1 \theta \exp\left(\frac{-\lambda_1 x}{1-\theta}\right) \int_0^x \exp\left(\frac{\lambda_1 t}{1-\theta} - \lambda_1 t\right) dt \\
&= \lambda_1 \theta \exp\left(\frac{-\lambda_1 x}{1-\theta}\right) \left[\frac{\exp\left(\frac{\lambda_1 \theta t}{1-\theta}\right)}{\lambda_1 \theta (1-\theta)} \right]_0^x \\
&= (1-\theta) \left[\exp\left(\frac{\lambda_1 \theta x}{1-\theta} - 1\right) \exp\left(\frac{-\lambda_1 x}{1-\theta}\right) \right] \\
&= (1-\theta) \left[\exp\left(\frac{-\lambda_1 x(1-\theta)}{1-\theta}\right) - \exp\left(\frac{-\lambda_1 x}{1-\theta}\right) \right] \\
&= (1-\theta) \left[\exp(-\lambda_1 x) - \exp\left(\frac{-\lambda_1 x}{(1-\theta)}\right) \right].
\end{aligned}$$

Par conséquent, nous avons

$$\begin{aligned}
P(X > x) &= (1-\theta) \exp\left(\frac{-\lambda_1 x}{1-\theta}\right) + (1-\theta) \left[\exp(-\lambda_1 x) - \exp\left(\frac{-\lambda_1 x}{(1-\theta)}\right) \right] + \theta e^{-\lambda_1 x} \\
&= e^{-\lambda_1 x}.
\end{aligned}$$

Ce qui montre que X suit une loi exponentielle de paramètre λ_1 . On peut montrer d'une manière analogue que Y suit une loi exponentielle de paramètre λ_2 , ce qui complète l'argument. \square

Nous remarquons que le modèle de Raftery décrit dans la définition 5.1.1 s'obtient comme cas particulier du modèle proposé introduit dans la définition 5.1.2, lorsque $\lambda_1 = \lambda_2 = \lambda$.

4.2.1 Structure de corrélation du modèle exponentiel proposée $\mathcal{E}(\theta, \lambda_1, \lambda_2)$

Dans cette partie, nous établissons l'expression de la corrélation associée à la famille des lois exponentielles $\mathcal{E}(\theta, \lambda_1, \lambda_2)$.

Proposition 4.2.2. *Pour tout $(X, Y) \in \mathcal{E}(\theta, \lambda_1, \lambda_2)$, nous avons*

$$\text{cor}(X, Y) = \theta(1 - \theta).$$

Démonstration. Pour ce faire, calculons d'abord la covariance entre X et Y . Comme, les variables aléatoires Z_1, Z_2, Z et J sont indépendantes, donc :

$$\begin{aligned}\text{cov}(X, Y) &= \text{cov}\left((1 - \theta)Z_1 + J\frac{Z}{\lambda_1}, (1 - \theta)Z_2 + J\frac{Z}{\lambda_2}\right) \\ &= \frac{1}{\lambda_1\lambda_2}\text{cov}(JZ, JZ) \\ &= \frac{1}{\lambda_1\lambda_2}\text{var}(JZ).\end{aligned}$$

Il reste donc à évaluer la variance de JZ . Par définition, nous avons

$$\text{var}(JZ) = E(JZ^2) - [E(JZ)]^2 = \theta(2 - \theta),$$

car

$$\begin{aligned}E((JZ)^2) &= E(JZ^2) = \theta E(Z^2) = 2\theta \\ [E(JZ)]^2 &= [\theta E(Z)]^2 = \theta^2.\end{aligned}$$

Par conséquent,

$$\text{cov}(X, Y) = \frac{\theta(2 - \theta)}{\lambda_1\lambda_2}.$$

Finalement,

$$\text{cor}(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)}\sqrt{\text{var}(Y)}} = \frac{\frac{\theta(2-\theta)}{\lambda_1\lambda_2}}{\frac{1}{\lambda_1\lambda_2}} = \theta(2 - \theta).$$

□

Nous constatons que le coefficient de corrélation de cette famille ne dépend que du paramètre de dépendance $\theta \in [0, 1]$. De plus, la corrélation s'annule lorsque $\theta = 0$, situation qui représente le cas d'indépendance. De même, le coefficient de corrélation atteint la valeur 1 lorsque $\theta = 1$, situation décrivant le cas où les variables aléatoires X et Y sont parfaitement dépendantes. Nous concluons que cette famille modélise tout le spectre des corrélations positives.

4.3 Copule de survie associée à la famille $\mathcal{E}(\theta, \lambda_1, \lambda_2)$

Nous notons par \bar{H} , \bar{F} et \bar{G} les fonction de survie respectivement des variables aléatoires (X, Y) , X et Y . Posons

$$V = \bar{F}(X) \quad \text{et} \quad W = \bar{G}(Y).$$

Pour faciliter le calcul de la copule de survie \hat{C} de (X, Y) . Remarquons d'abord que cette dernière coïncide avec la fonction de répartition de (U, V) . En effet,

$$\begin{aligned} P(V \leq u, W \leq v) &= P(\bar{F}(X) \leq u, \bar{G}(Y) \leq v) \\ &= P(X \geq \bar{F}^{-1}(u), Y \geq \bar{G}^{-1}(v)) \\ &= \bar{H}(X \geq \bar{F}^{-1}(u), Y \geq \bar{G}^{-1}(v)) \\ &= \hat{C}(u, v). \end{aligned}$$

Puisque les variables aléatoires Z_1, Z_2 et Z sont des variables telles que $Z_1 \sim \text{Exp}(\lambda_1)$, $Z_2 \sim \text{Exp}(\lambda_2)$ et $Z \sim \text{Exp}(1)$. Donc, il existe des variables aléatoires U_1, U_2 et U uniformément distribuées sur l'intervalle $[0, 1]$ telles que

$$Z_1 = -\frac{1}{\lambda_1} \ln(U_1), \quad Z_2 = -\frac{1}{\lambda_2} \ln(U_2), \quad Z = -\ln(U).$$

Par conséquent,

$$\begin{aligned} V = \bar{F}(X) &= \exp \left[-\lambda_1 \left(-\frac{(1-\theta)}{\lambda_1} \ln(U_1) - J \frac{1}{\lambda_1} \ln(U) \right) \right] \\ &= U_1^{1-\theta} U^J. \end{aligned}$$

De même,

$$\begin{aligned} W = \bar{G}(Y) &= \exp \left[-\lambda_2 \left(-\frac{(1-\theta)}{\lambda_2} \ln(U_2) - J \frac{1}{\lambda_2} \ln(U) \right) \right] \\ &= U_2^{1-\theta} U^J. \end{aligned}$$

Proposition 4.3.1. *La copule de survie de $(X, Y) \in \mathcal{E}(\theta, \lambda_1, \lambda_2)$ est donnée par*

$$\hat{C}(u, v) = \min(u, v) + \frac{1+\theta}{1-\theta} u^{\frac{1}{1-\theta}} v^{\frac{1}{1-\theta}} \left[1 - \max(u, v)^{-\frac{1+\theta}{1-\theta}} \right].$$

Démonstration. D'après ce qui précède, la copule recherchée est la fonction de répartition du couple aléatoire

$$V = U_1^{1-\theta} U^J, \quad W = U_2^{1-\theta} U^J. \quad (4.1)$$

En conditionnant par la variable aléatoire de Bernoulli J de paramètre θ indépendante des variables aléatoires U_1, U_2, U , on obtient que

$$\hat{C}(u, v) = (1-\theta)P(U_1^{1-\theta} \leq u, U_2^{1-\theta} \leq v) + \theta P(U_1^{1-\theta} U^J \leq u, U_2^{1-\theta} U^J \leq v)$$

Le premier terme du membre droit de l'équation ci-dessus se calcule simplement par

$$T_1 = (1-\theta)P(U_1^{1-\theta} \leq u, U_2^{1-\theta} \leq v) = (1-\theta)u^{\frac{1}{1-\theta}}v^{\frac{1}{1-\theta}}$$

Remarquons maintenant que le terme $\theta P(U_1^{1-\theta}U \leq u, U_2^{1-\theta}U \leq v)$ est échangeable en (u, v) , car les variables aléatoires U_1, U_2, U sont indépendantes et identiquement distribuées, ce qui permet d'écrire :

$$\theta P(U_1^{1-\theta}U \leq u, U_2^{1-\theta}U \leq v) = \theta P(U_1^{1-\theta}U \leq \min(u, v), U_2^{1-\theta}U \leq \max(u, v)).$$

En conditionnant à nouveau par la variable aléatoire U , il vient que

$$\begin{aligned} & \theta P(U_1^{1-\theta}U \leq \min(u, v), U_2^{1-\theta}U \leq \max(u, v)) \\ = & \theta \int_0^1 P\left(U_1^{1-\theta} \leq \frac{\min(u, v)}{t}, U_2^{1-\theta} \leq \frac{\max(u, v)}{t}\right) dt \\ = & \theta \int_0^{\min(u, v)} dt \\ + & \theta \int_{\min(u, v)}^{\max(u, v)} P\left(v_1^{1-\theta} \leq \frac{\min(u, v)}{t}\right) dt \\ + & \theta \int_{\max(u, v)}^1 \left(\frac{\min(u, v)}{t}\right)^{\frac{1}{1-\theta}} \left(\frac{\max(u, v)}{t}\right)^{\frac{1}{1-\theta}} dt. \end{aligned}$$

Notons que

$$T_2 = \theta \int_0^{\min(u, v)} dt = \theta \min(u, v)$$

D'une manière analogue, nous obtenons

$$\begin{aligned} T_3 &= \theta \int_{\min(u, v)}^{\max(u, v)} P\left(v_1^{1-\theta} \leq \frac{\min(u, v)}{t}\right) dt \\ &= \theta \int_{\min(u, v)}^{\max(u, v)} [\min(u, v)]^{\frac{1}{1-\theta}} t^{-\frac{1}{1-\theta}} dt \\ &= \theta \min(u, v)^{\frac{1}{1-\theta}} \int_{\min(u, v)}^{\max(u, v)} t^{-\frac{1}{1-\theta}} dt \\ &= \theta \min(u, v)^{\frac{1}{1-\theta}} \left[\frac{-(1-\theta)t^{-\frac{\theta}{1-\theta}}}{\theta} \right]_{\min(u, v)}^{\max(u, v)} \\ &= \min(u, v)^{\frac{1}{1-\theta}} (1-\theta) [\min(u, v)^{-\frac{\theta}{1-\theta}} - \max(u, v)^{-\frac{\theta}{1-\theta}}] \\ &= (1-\theta) \min(u, v) - (1-\theta) \min(u, v)^{\frac{1}{1-\theta}} \max(u, v)^{-\frac{\theta}{1-\theta}}, \end{aligned}$$

$$\begin{aligned} T_4 &= \theta \int_{\max(u, v)}^1 \left(\frac{\min(u, v)}{t}\right)^{\frac{1}{1-\theta}} \left(\frac{\max(u, v)}{t}\right)^{\frac{1}{1-\theta}} dt = \theta u^{\frac{1}{1-\theta}} v^{\frac{1}{1-\theta}} \int_{\max(u, v)}^1 t^{-\frac{2}{1-\theta}} dt \\ &= \frac{\theta(1-\theta)}{1+\theta} \left[\max(u, v)^{-\frac{(1+\theta)}{1-\theta}} - 1 \right]. \end{aligned}$$

Finalement, en sommant les 4 termes T_1, T_2, T_3 et T_4 , nous concluons que

$$\begin{aligned}
 C(u, v) &= (1 - \theta)u^{\frac{1}{1-\theta}}v^{\frac{1}{1-\theta}} + \min(u, v) \\
 &- (1 - \theta)\min(u, v)^{\frac{1}{1-\theta}}\max(u, v)^{\frac{-\theta}{1-\theta}} \\
 &+ \frac{\theta(1 - \theta)}{1 + \theta}[\max(u, v)^{\frac{-(1+\theta)}{1-\theta}} - 1] \\
 &= \min(u, v) \\
 &+ \frac{1 + \theta}{1 - \theta}u^{\frac{1}{1-\theta}}v^{\frac{1}{1-\theta}} \left[1 + \theta - (1 + \theta)\min(u, v)^{\frac{1}{1-\theta}}\max(u, v)^{\frac{-\theta}{1-\theta}}u^{-\frac{1}{1-\theta}}v^{-\frac{1}{1-\theta}} \right] \\
 &+ \frac{1 + \theta}{1 - \theta}u^{\frac{1}{1-\theta}}v^{\frac{1}{1-\theta}} \left[\theta(\max(u, v)^{\frac{-1+\theta}{1-\theta}} - 1) \right].
 \end{aligned} \tag{4.2}$$

Remarquons que $\min(u, v)^{\frac{1}{1-\theta}}\max(u, v)^{\frac{-\theta}{1-\theta}}u^{-\frac{1}{1-\theta}}v^{-\frac{1}{1-\theta}}$ coïncide avec

$$\min(u, v)^{\frac{1}{1-\theta}}\max(u, v)^{\frac{-\theta}{1-\theta}}\min(u, v)^{-\frac{1}{1-\theta}}\max(u, v)^{-\frac{1}{1-\theta}} = \max(u, v)^{-\frac{1+\theta}{1-\theta}}. \tag{4.3}$$

En insérant (4.3) dans (4.2), il vient à partir de (4.2) que

$$\hat{C}(u, v) = \min(u, v) + \frac{1 + \theta}{1 - \theta}u^{\frac{1}{1-\theta}}v^{\frac{1}{1-\theta}} \left[1 - \max(u, v)^{-\frac{1+\theta}{1-\theta}} \right].$$

□

4.3.1 Algorithme de simulation de la copule de survie de la famille $\mathcal{E}(\theta, \lambda_1, \lambda_2)$

La représentation stochastique la copule de survie associée à $\mathcal{E}(\theta, \lambda_1, \lambda_2)$ décrit par les équations (4.1) permet de générer facilement des données issues de cette copule. En effet, l'algorithme ci-dessous indique les étapes principales de simulation de données provenant de cette copule :

1. Fixons $\theta \in (0, 1)$
2. Générer des aléas indépendantes u_1, u_2, u de loi uniforme sur $[0, 1]$
3. Générer aléas j de la de bernoulli de paramètre θ
4. Posons $v = u_1^{1-\theta}u^j$ et $w = u_2^{1-\theta}u^j$
5. Le couple désiré sera (v, w)

Ci-dessous, quelques graphiques illustrant le comportement de données générées à partir de la copules survie de la famille $\mathcal{E}(\theta, \lambda_1, \lambda_2)$.

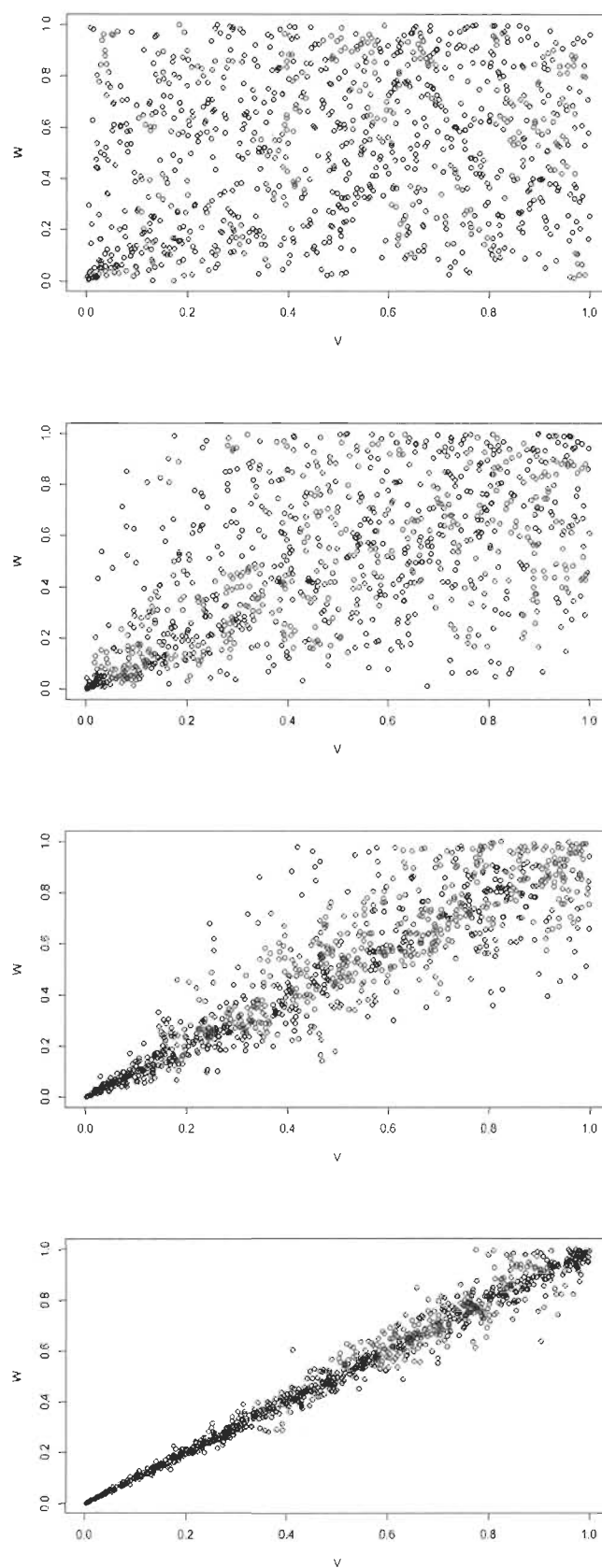


FIGURE 4.1 – Simulation de la copule de Raftery avec $\theta = 0.3, 0.5, 0.8$ et 0.95

4.4 Mesures de concordances correspondantes à la famille $\mathcal{E}(\theta, \lambda_1, \lambda_2)$

4.4.1 Tau de Kendall de la famille $\mathcal{E}(\theta, \lambda_1, \lambda_2)$

Proposition 4.4.1. *Le tau de Kendall associé à la famille $\mathcal{E}(\theta, \lambda_1, \lambda_2)$ est exprimé par :*

$$\tau_\theta = \frac{2\theta}{3 - \theta}$$

Démonstration. Pour calculer le tau de Kendall, nous devons d'abord déterminer la densité de la copule de survie. Cette dernière s'obtient facilement en dérivant successivement par rapport à u et v la copule de survie.

$$C(u, v) = \begin{cases} u + \frac{1 - \theta}{1 + \theta} (uv)^{\frac{1}{1-\theta}} \left(1 - v^{-\frac{1+\theta}{1-\theta}}\right), & u \leq v, \\ v + \frac{1 - \theta}{1 + \theta} (uv)^{\frac{1}{1-\theta}} \left(1 - u^{-\frac{1+\theta}{1-\theta}}\right), & u > v. \end{cases}$$

Ainsi, pour $u < v$ calculons $\frac{\partial C}{\partial u \partial v}$, ainsi nous avons

$$\begin{aligned} \frac{\partial C}{\partial u} &= \frac{\partial}{\partial u} \left\{ u + \frac{1 - \theta}{1 + \theta} (uv)^{\frac{1}{1-\theta}} \left(1 - v^{-\frac{1+\theta}{1-\theta}}\right) \right\} \\ &= 1 + \frac{1 - \theta}{1 + \theta} v^{\frac{1}{1-\theta}} \left(1 - v^{-\frac{1+\theta}{1-\theta}}\right), \frac{\partial}{\partial u} u^{\frac{1}{1-\theta}} \\ &= 1 + \frac{1}{1 + \theta} u^{\frac{\theta}{1-\theta}} v^{\frac{1}{1-\theta}} \left(1 - v^{-\frac{1+\theta}{1-\theta}}\right), \end{aligned}$$

En dérivant par rapport à v , on obtient

$$\frac{\partial^2 C}{\partial u \partial v} = \frac{1}{1 - \theta^2} u^{\frac{\theta}{1-\theta}} (v^{\frac{\theta}{1-\theta}} + \theta v^{\frac{-1}{1-\theta}}) \quad (4.4)$$

De façon analogue, on obtient pour le cas $u > v$,

$$\frac{\partial^2 C}{\partial u \partial v} = \frac{1}{1 - \theta^2} v^{\frac{\theta}{1-\theta}} (u^{\frac{\theta}{1-\theta}} + \theta u^{\frac{-1}{1-\theta}}) \quad (4.5)$$

Rappelons que le tau de Kendall s'exprime en fonction de la copule C comme suit :

$$\begin{aligned} \tau &= 4 \int_0^1 \int_0^1 C(u, v) \frac{\partial^2 C}{\partial u \partial v} du dv - 1 \\ &= 4 \int_0^1 \left(\int_0^v C(u, v) \frac{\partial^2 C}{\partial u \partial v} du + \int_v^1 C(u, v) \frac{\partial^2 C}{\partial u \partial v} du \right) dv - 1. \end{aligned} \quad (4.6)$$

En insérant les expressions (4.4) et (4.5) dans (4.6), et en intégrant successivement par rapport à u et v , il vient après simplifications que

$$\tau = \frac{2\theta}{3 - \theta},$$

ce qui conduit au résultat escompté.

□

4.4.2 Rho de Spearman de la famille $\mathcal{E}(\theta, \lambda_1, \lambda_2)$

Proposition 4.4.2. *Le rho de Spearman associé à la famille $\mathcal{E}(\theta, \lambda_1, \lambda_2)$ est exprimé par :*

$$\rho_\theta = \frac{\theta(4 - 3\theta)}{(2 - \theta)^2}.$$

Démonstration. Rappelons que le rho de Spearman s'exprime en fonction de la copule de survie de la famille $\mathcal{E}(\theta, \lambda_1, \lambda_2)$ à travers les variables V et W définies par les relations. Ainsi, nous avons

$$\rho_\theta = 12E(VW) - 3.$$

En utilisant le fait que les variables aléatoires U_1, U_2 et U sont indépendantes, il vient que

$$\begin{aligned} E(VW) &= (1 - \theta)E(U_1^{1-\theta}U_2^{1-\theta}) + \theta E(U_1^{1-\theta}U_2^{1-\theta}U^2) \\ &= (1 - \theta)E(U_1^{1-\theta})E(U_2^{1-\theta}) + \theta E(U_1^{1-\theta})E(U_2^{1-\theta})E(U^2) \end{aligned}$$

avec

$$E(U_1^{1-\theta}) = \int_0^1 u^{1-\theta} du = \frac{u^{2-\theta}}{2-\theta} \Big|_0^1 = \frac{1}{2-\theta} = E(U_2^{1-\theta}).$$

De plus, on peut montrer facilement que $E(U^2) = \frac{1}{3}$. Par conséquent :

$$\begin{aligned} E(VW) &= (1 - \theta)E(U_1^{1-\theta}U_2^{1-\theta}) + \theta E(U_1^{1-\theta}U_2^{1-\theta}U^2) \\ &= \frac{1 - \theta}{(2 - \theta)^2} + \frac{\theta}{3(2 - \theta)^2} \\ &= \frac{3 - 2\theta}{3(2 - \theta)^2} \end{aligned}$$

Par la suite,

$$\begin{aligned}
 \rho_\theta &= 12E(UVW) - 3 \\
 &= \frac{4(3 - 2\theta)}{(2 - \theta)^2} - 3 \\
 &= \frac{12 - 8\theta - 3(4 - 4\theta + \theta^2)}{(2 - \theta)^2} \\
 &= \frac{4\theta - 3\theta^2}{(2 - \theta)^2} \\
 &= \frac{\theta(4 - 3\theta)}{(2 - \theta)^2}.
 \end{aligned}$$

□

Nous observons que ρ_θ est une fonction croissante de paramètre de dépendance $\theta \in [0, 1]$. De plus, $\rho_\theta = 0$ si $\theta = 0$, et $\rho_\theta = 1$ si $\theta = 1$.

4.5 Étude de l'estimateur des moments du paramètre de dépendance de la copule de Raftery

4.5.1 Estimateur du paramètre de dépendance à partir de la corrélation

Proposition 4.5.1. *L'estimateur de moment du paramètre de dépendance θ , basé sur la corrélation de Pearson, est donné par*

$$\hat{\theta}_{cp} = 1 - \sqrt{1 - r}$$

Démonstration. Rappelons que la structure de corrélation du modèle exponentiel $\mathcal{E}(\theta, \lambda_1, \lambda_2)$ ne dépend que du paramètre de dépendance $\theta \in [0, 1]$. Ainsi nous avons

$$r = \theta(2 - \theta) = 2\theta - \theta^2.$$

Ainsi, l'estimateur $\hat{\theta}_{cp}$ de θ sera la solution positive de l'équation du second degré

$$\theta^2 - 2\theta + r = 0,$$

à savoir

$$\hat{\theta}_{cp} = 1 - \sqrt{1 - r}$$

qui est bel bien un élément de l'intervalle $\theta \in [0, 1]$.

□

4.5.2 Estimateur du paramètre de dépendance à partir du rho de Spearman

Proposition 4.5.2. *L'estimateur de moment du paramètre de dépendance θ basé sur la corrélation de Spearman, est donné par*

$$\hat{\theta}_{sp} = \frac{2(1 + r) - 2\sqrt{1 - r}}{3 + r}.$$

Démonstration. Le procédé est très similaire au précédant pour la corrélation de Person. En effet, l'estimateur $\hat{\theta}_{sp}$ de θ sera la solution de l'équation

$$r = \frac{\theta(4 - 3\theta)}{(2 - \theta)^2},$$

Ce qui est équivalent à résoudre l'équation de second degré suivante :

$$(3 + r)\theta^2 - 4(1 + r)\theta + 4r = 0.$$

La solution positive de cette dernière fournit l'estimateur désiré à savoir

$$\hat{\theta}_{sp} = \frac{2(1 + r) - 2\sqrt{1 - r}}{3 + r}$$

d'où le résultat.

□

4.5.3 Estimateur du paramètre de dépendance à partir du tau de Kendall

Proposition 4.5.3. *L'estimateur de moment du paramètre de dépendance θ basé sur le tau de Kendall, est donné par*

$$\hat{\theta}_k = \frac{3\hat{\tau}}{(2 + \hat{\tau})}.$$

Démonstration. L'estimateur des moments basé sur le tau de Kendall se déduit aisément à partir de l'expression du tau de Kendall en fonction θ . Par conséquent,

$$\begin{aligned}
 \hat{\tau} = \frac{2\hat{\theta}}{(3 - \hat{\theta})} &\iff (3 + r)\theta^2 - 4(1 + r)\theta + 4r = 0 \\
 &\iff 3\hat{\tau} = 2\hat{\theta} + \hat{\theta}\hat{\tau} \\
 &\iff 3\hat{\tau} = \theta(2 + \hat{\tau}) \\
 &\iff \hat{\theta}_k = \frac{3\hat{\tau}}{(2 + \hat{\tau})}.
 \end{aligned}$$

Donc, nous avons trois estimateurs du paramètre de dépendance de la copule de Raftery. Il sera intéressant de comparer les estimateurs par la suite.

□

4.6 Comparaison des estimateurs par simulation

Pour conclure l'étude du nouveau modèle exponentiel bidimensionnel de type Raftery, effectuons une comparaison de l'étude de l'estimateur des moments du paramètre de dépendance de la copule de Raftery. Effectuons un tableau avec les estimateurs de θ . Pour différentes tailles d'échantillons, on obtient les résultats ci-dessous.

Estimateur	n=20	n=30	n=50
$\hat{\theta}_{sp}$	0.27583	0.27727	0.301274
$EQM(\hat{\theta}_{sp})$	8.069×10^{-5}	0.00051	1.624×10^{-6}
$\hat{\theta}_{cp}$	0.17120	0.169714	0.18337
$EQM(\hat{\theta}_{cp})$	0.0166	0.0170	0.013601

FIGURE 4.2 – Simulation des estimateurs de la copule de Raftery avec $\theta = 0.3$

Estimateur	n=100	n=150	n=200
$\hat{\theta}_{sp}$	0.28021	0.28481	0.28491
$EQM(\hat{\theta}_{sp})$	0.00039	0.00023	0.00023
$\hat{\theta}_{cp}$	0.18687	0.16943	0.17195
$EQM(\hat{\theta}_{cp})$	0.01280	0.01705	0.01640

FIGURE 4.3 – Simulation des estimateurs de la copule de Raftery avec $\theta = 0.3$

Estimateur	n=20	n=30	n=50
$\hat{\theta}_{sp}$	0.46770	0.47693	0.48048
$EQM(\hat{\theta}_{sp})$	0.00104	0.00053	0.00038
$\hat{\theta}_{cp}$	0.35352	0.31844	0.33607
$EQM(\hat{\theta}_{cp})$	0.02146	0.03297	0.02687

FIGURE 4.4 – Simulation des estimateurs de la copule de Raftery avec $\theta = 0.5$

Estimateur	n=100	n=150	n=200
$\hat{\theta}_{sp}$	0.50556	0.50600	0.48633
$EQM(\hat{\theta}_{sp})$	0.00039	3.553×10^{-5}	0.00019
$\hat{\theta}_{cp}$	0.32954	0.33502	0.33471
$EQM(\hat{\theta}_{cp})$	0.02905	0.02722	0.02732

FIGURE 4.5 – Simulation des estimateurs de la copule de Raftery avec $\theta = 0.5$

$$EQM(\hat{\theta}_{sp}) = \frac{1}{k} \sum_{i=1}^k (\hat{\theta}_{sp} - \theta)^2$$

$$EQM(\hat{\theta}_{cp}) = \frac{1}{k} \sum_{i=1}^k (\hat{\theta}_{cp} - \theta)^2$$

Pour vérifier la qualité des estimateurs, considérons l'erreur quadratique moyenne des estimateurs simulés pour $k = 100$ répétitions. La même simulation a été effectuée pour $k = 1000$ et les résultats sont très similaires. Il est important de noter que peu importe la taille d'échantillon, l'estimateur du paramètre de dépendance par la corrélation de Person reste imprécis. On favorise l'estimateur par la corrélation de Spearman.

4.7 Fonction de survie de la famille exponentielle bivariée $\mathcal{E}(\theta, \lambda_1, \lambda_2)$

Dans cette section nous pouvons établir l'expression de la fonction de survie \bar{H} de la famille exponentielle bivariée proposée $\mathcal{E}(\theta, \lambda_1, \lambda_2)$. Celle-ci peut être obtenue à partir de la copule de survie établie dans la proposition 4.3.1 comme montre le résultat suivant.

Proposition 4.7.1. *La fonction de survie associée à la famille $\mathcal{E}(\theta, \lambda_1, \lambda_2)$ est donnée par*

$$\begin{aligned} \bar{H}(x, y) &= \min(\exp(-\lambda_1 x), \exp(-\lambda_2 y)) \\ &+ \frac{1+\theta}{1-\theta} \exp\left(-\frac{\lambda_1 x}{1-\theta}\right) \exp\left(-\frac{\lambda_2 y}{1-\theta}\right) \left[1 - \{\exp[-\min(-\lambda_1 x, -\lambda_2 y)]\}^{-\frac{1+\theta}{1-\theta}}\right]. \end{aligned}$$

Démonstration. D'après le théorème de Sklar (1958), nous avons

$$\bar{H}(x, y) = \hat{C}(\bar{F}(x), \bar{G}(y)) = \hat{C}(\exp(-\lambda_1 x), \exp(-\lambda_2 y)).$$

En utilisant le fait que

$$\max[\exp(-\lambda_1 x), \exp(-\lambda_2 y)] = \exp[-\min(-\lambda_1 x, -\lambda_2 y)]$$

et en substituant respectivement u et v par $\exp(-\lambda_1 x)$ et $\exp(-\lambda_2 y)$ dans l'expression de \hat{C} explicité dans la proposition 4.3.1, nous déduisons la formule de \bar{H} désirée. \square

4.7.1 Algorithme de simulation de la fonction de survie de la famille $\mathcal{E}(\theta, \lambda_1, \lambda_2)$

La simulation de données issues de la famille $\mathcal{E}(\theta, \lambda_1, \lambda_2)$ peut être effectuée via la représentation stochastique d'un couple de variables aléatoires $(X, Y) \in \mathcal{E}(\theta, \lambda_1, \lambda_2)$ donnée par

$$\begin{aligned} X &= (1 - \theta)Z_1 + J\frac{Z}{\lambda_1}, \\ Y &= (1 - \theta)Z_2 + J\frac{Z}{\lambda_2}, \end{aligned}$$

où Z_1, Z_2, Z des variables aléatoires indépendantes telles que $Z_1 \sim \text{Exp}(\lambda_1)$, $Z_2 \sim \text{Exp}(\lambda_2)$ et $Z \sim \text{Exp}(1)$ et J qui suit une loi de Bernoulli de paramètre θ qui est indépendante de Z_1, Z_2 et Z . Ainsi l'algorithme ci-dessous indique les étapes principales de simulation de données provenant de la famille $\mathcal{E}(\theta, \lambda_1, \lambda_2)$:

1. Fixons λ_1, λ_2 et θ
2. Générer des aléas indépendantes u_1, u_2, u de loi uniforme sur $[0, 1]$
3. posons $z_1 = -\frac{1}{\lambda_1} \ln(u_1)$, $z_2 = -\frac{1}{\lambda_2} \ln(u_2)$ et $z = -\ln(u)$
4. Générer aléas j de la de bernoulli de paramètre θ
5. Posons

$$x = (1 - \theta)z_1 + j\frac{z}{\lambda_1}, \quad y = (1 - \theta)z_2 + j\frac{z}{\lambda_2}.$$

6. Le couple désiré sera (x, y)

Une autre manière de générer des données de la famille $\mathcal{E}(\theta, \lambda_1, \lambda_2)$ consiste à simuler des observations (u, v) de la copule de survie \hat{C} , puis poser $x = \exp(-\lambda_1 u)$ et $y = \exp(-\lambda_2 v)$. Ainsi le couple (x, y) provient de la fonction de survie \bar{H} .

Il est intéressant d'observer la tendance de la survie du modèle sous plusieurs valeurs distinctes de θ . Alors, les graphiques ci-dessous représentera les tendances pour $n = 500$ observations. On choisit $n = 500$ comme taille d'échantillon car il y a une grande concentration sur une ligne et quelques données éparses.

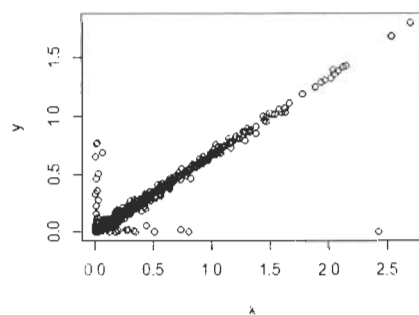
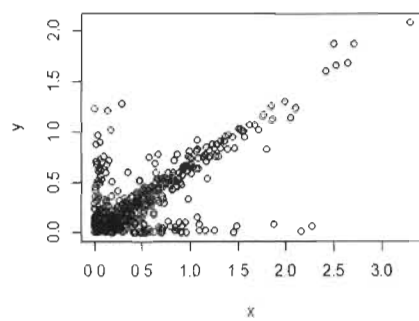
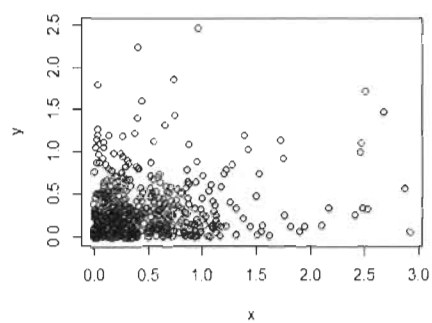
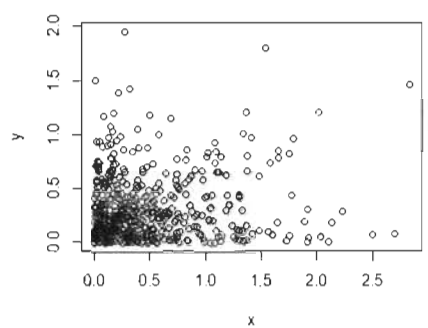


FIGURE 4.6 – Simulation de la fonction de survie du modèle de Raftery avec $\theta = 0.3, 0.5, 0.8$ et 0.95

Chapitre 5

Extension du modèle de Raftery au cas de dépendance négative

Le modèle de Raftery (6) (1984) ainsi que l'extension proposés dans le chapitre précédent modélisent uniquement les paires aléatoires exponentielles, dont les lois marginales sont positivement dépendantes. Dans ce chapitre, nous proposons une loi exponentielle bidimensionnelle de type Raftery spécifique pour le cas où les variables marginales sont négativement dépendantes. Ce modèle utilise la notion de contre-monotone pour induire la dépendance négative entre les composantes de la distribution exponentielle bidimensionnelle.

Dans tout ce chapitre, nous notons respectivement par F_1 et F_2 les fonctions de répartition des variables aléatoires X_1 et X_2 , et par \bar{F}_1 et \bar{F}_2 leurs fonctions de survie respectives. Les fonctions de répartition et de survie jointes de (X_1, X_2) seront notées respectivement H et \bar{H} .

5.1 Modèle exponentiel bidimensionnel négatif de type Raftery

Le modèle proposé sera la notion de contre-monotone, qui caractérise la dépendance négative parfaite.

Définition 5.1.1. Soient X_1 et X_2 , deux variables aléatoires. On dit que les variables

aléatoire X_1 et X_2 sont contre-comonotones si la fonction de répartition jointe (X_1, X_2) coïncide avec la borne de Fréchet supérieure, à savoir :

$$H(x, y) = \sup(F_1(x) + F_2(y) - 1, 0), \quad \text{pour tout } (x, y) \in \mathbb{R}^2$$

Notons que si deux variables aléatoires sont contre-comonotones, alors il existe une variable aléatoire U uniformément distribuée sur l'intervalle $[0, 1]$ tels que

$$X_1 = F_1^{-1}(U) \quad \text{et} \quad X_2 = F_2^{-1}(1 - U).$$

Ceci montre que les variables aléatoires contre-comonotones sont parfaitement négativement dépendantes. La loi exponentielle bivariée proposée est construite en liant des composantes exponentielles indépendantes par des variables aléatoires contre-comonotones.

Définition 5.1.2. Soient Z_1, Z_2, Z_3, Z_4 , des variables aléatoires telles que

$$Z_1, Z_3 \sim \text{Exp}(\lambda_1) \quad \text{et} \quad Z_2, Z_4 \sim \text{Exp}(\lambda_2).$$

De plus, nous supposons que :

- Z_1, Z_2, Z_3 sont indépendantes ;
- Z_1, Z_2, Z_4 sont indépendantes ;
- Z_3, Z_4 sont contre-comonotone.

Soient J_1 et J_2 , des variables aléatoires de Bernoulli indépendantes et identiquement distribuées de paramètre θ telles que (J_1, J_2) et (Z_1, Z_2, Z_3, Z_4) . Le couple de variables aléatoires défini par

$$\begin{aligned} X_1 &= (1 - \theta)Z_1 + J_1Z_3, \\ X_2 &= (1 - \theta)Z_2 + J_1Z_4, \end{aligned}$$

suit une loi exponentielle bidimensionnelle de type Raftery, notée $\mathcal{E}^-(\lambda_1, \lambda_2, \theta)$. De plus, les lois marginales X_1 et X_2 ne sont pas affectées par le paramètre de dépendance θ , et suivent respectivement des lois exponentielles de paramètres λ_1 et λ_2 .

Il est clair que les paires aléatoires éléments de $\mathcal{E}^-(\lambda_1, \lambda_2, \theta)$ modélisent uniquement la dépendance négative.

Nous proposons ci-dessous une autre formulation plus intéressante de la famille de distributions bivariées $\mathcal{E}^-(\lambda_1, \lambda_2, \theta)$.

Proposition 5.1.1. *Soient U_1, U_2 et V des variables aléatoires indépendantes uniformément distribuées sur l'intervalle $[0, 1]$. Soient J_1 et J_2 , des variables aléatoires de Bernoulli indépendantes et identiquement distribuées de paramètre θ . Les vecteurs aléatoires (J_1, J_2) et (U_1, U_2, V) sont supposés indépendants. Le couple de variables aléatoires défini par*

$$X_1 = -\frac{(1-\theta)}{\lambda_1} \ln(U_1) - \frac{J_1}{\lambda_1} \ln(V); \quad (5.1)$$

$$X_2 = -\frac{(1-\theta)}{\lambda_2} \ln(U_2) - \frac{J_2}{\lambda_2} \ln(1-V). \quad (5.2)$$

suit une loi exponentielle bidimensionnelle de type Raftery, notée $\mathcal{E}^-(\lambda_1, \lambda_2, \theta)$.

Démonstration. Ceci découle du fait pour toute variable aléatoire exponentielle X de paramètre λ , il existe une variable U des variables aléatoires uniformément distribuées sur $[0, 1]$ telle

$$X = -\frac{1}{\lambda} \ln(U).$$

□

En fixant $\lambda_1 > 0$ et $\lambda_2 > 0$, soit $\Gamma(F_1, F_2)$ l'espace des modèles exponentiels bidimensionnels de lois marginales exponentielles de paramètres $\lambda_1 > 0$ et $\lambda_2 > 0$. Il est bien connu que le support des corrélations de Pearson dans l'espace $\Gamma(F_{\lambda_1}, F_{\lambda_2})$ est l'intervalle $[\rho_{\min}, 1]$. Notons que la corrélation minimale est calculée à l'aide de la borne inférieure de Fréchet $H_{\min}(x_1, x_2) = \sup(F_1(x_1) + F_2(x_2) - 1, 0)$. Cette borne inférieure de la corrélation est explicitement donnée par formule suivante.

$$\rho_{\min} = \int_0^1 \ln(u) \ln(1-u) du - 1 = 1 - \pi^2/6.$$

5.2 Structure de corrélation de la famille $\mathcal{E}^-(\lambda_1, \lambda_2, \theta)$

Le but de cette section est d'examiner la structure de corrélation de la famille $\mathcal{E}^-(\lambda_1, \lambda_2, \theta)$. Nous montrons dans la proposition suivante que cette structure de corrélation est intimement liée au paramètre de dépendance θ et la corrélation minimale ρ_{\min} .

Théorème 5.2.1.

$$r_\theta = \text{cor}(X_1, X_2) = \left(1 - \frac{\pi^2}{6}\right) \theta^2.$$

Démonstration. Puisque les variables aléatoires U_1, U_2, U impliquées dans la définition de (X_1, X_2) décrite par les équations (5.1) et (5.2) sont indépendantes, donc

$$\begin{aligned}
 \text{cov}(X_1, X_2) &= \frac{1}{\lambda_1 \lambda_2} \text{cov}(J_1 \ln(V), J_2 \ln(1 - V)) \\
 &= \frac{1}{\lambda_1 \lambda_2} [E \{J_1 J_2 \ln(V) \ln(1 - V)\} - E \{J_1 \ln(V)\} E \{J_2 \ln(1 - V)\}] \\
 &= \frac{\theta^2}{\lambda_1 \lambda_2} [E \{\ln(V) \ln(1 - V)\} - E \{\ln(V)\} E \{\ln(1 - V)\}] \\
 &= \frac{\theta^2}{\lambda_1 \lambda_2} [E \{\ln(V) \ln(1 - V)\} - 1]
 \end{aligned}$$

ce qui implique que

$$\text{cor}(X_1, X_2) = \theta^2 [E \{\ln(V) \ln(1 - V)\} - 1] = \theta^2 \rho_{\min} = \left(1 - \frac{\pi^2}{6}\right) \theta^2.$$

□

Observons que la corrélation r_θ est une fonction décroissante de θ avec $r_\theta = 0$ si $\theta = 0$ et $r_\theta = \rho_{\min}$ lorsque $\theta = 1$. Ceci illustre bien la structure de dépendance négative sous-jacente au $\mathcal{E}^-(\lambda_1, \lambda_2, \theta)$.

5.3 Copule de survie

Le but de cette section est d'établir la copule de survie correspondant à la famille de lois exponentielles bivariées $\mathcal{E}^-(\lambda_1, \lambda_2, \theta)$.

Proposition 5.3.1. *La copule de survie associée à la famille $\mathcal{E}^-(\lambda_1, \lambda_2, \theta)$ coïncide avec la fonction de répartition de couple aléatoire*

$$\tilde{U}_1 = U_1^{1-\theta} V^{J_1} \quad \text{et} \quad \tilde{U}_2 = U_2^{1-\theta} (1 - V)^{J_2}. \quad (5.3)$$

où U_1, U_2, V sont des variables aléatoires indépendantes uniformément distribuées sur l'intervalle $[0, 1]$ et J_1 et J_2 sont des variables aléatoires de Bernoulli indépendantes et identiquement distribuées de paramètre θ avec (J_1, J_2) et (U_1, U_2, V) sont supposés indépendants.

Démonstration. La copule de survie de $(X_1, X_2) \in \mathcal{E}^-(\lambda_1, \lambda_2, \theta)$ est donnée par la fonction de répartition jointe de $(\tilde{U}_1, \tilde{U}_2) = (\bar{F}_1(X_1), \bar{F}_2(X_2))$. D'après (5.1) et (5.2), nous

avons

$$\begin{aligned}\tilde{U}_1 &= \bar{F}_1(X_1) = \exp(-\lambda_1 X_1) = U_1^{1-\theta} V^{J_1} \\ \tilde{U}_2 &= \bar{F}_2(X_2) = \exp(-\lambda_2 X_2) = U_2^{1-\theta} (1-V)^{J_2}.\end{aligned}$$

□

Le théorème ci-dessous donne l'expression analytique de la copule de survie de la famille de distribution proposée.

Théorème 5.3.1. *La copule correspondant au modèle $\mathcal{E}^-(\lambda_1, \lambda_2, \theta)$ est*

si $u_1 > 1 - u_2$,

$$\begin{aligned}C(u_1, u_2) &= (1-\theta) \left\{ \left(1 - u_1^{\frac{1}{1-\theta}}\right) \left(1 - u_2^{\frac{1}{1-\theta}}\right) + u_1 u_2 \left[1 - \left(1 - u_1^{\frac{\theta}{1-\theta}}\right) \left(1 - u_2^{\frac{\theta}{1-\theta}}\right)\right] \right\} \\ &\quad + \theta(u_1 + u_2 - 1),\end{aligned}$$

si $u_1 \leq 1 - u_2$,

$$\begin{aligned}C(u_1, u_2) &= (1-\theta) \left\{ \left(1 - u_1^{\frac{1}{1-\theta}}\right) \left(1 - u_2^{\frac{1}{1-\theta}}\right) + u_1 u_2 \left[1 - \left(1 - u_1^{\frac{\theta}{1-\theta}}\right) \left(1 - u_2^{\frac{\theta}{1-\theta}}\right)\right] \right\} \\ &\quad + \theta(1-\theta) \left[u_2^{\frac{1}{1-\theta}} (1-u_1)^{-\frac{\theta}{1-\theta}} + u_1^{\frac{1}{1-\theta}} (1-u_2)^{-\frac{\theta}{1-\theta}} - 1 \right] + \theta^2 B_\theta(u_1, 1-u_2),\end{aligned}$$

où B_θ désigne la fonction définie par

$$B_\theta(a, b) = a^{\frac{1}{1-\theta}} (1-b)^{\frac{1}{1-\theta}} \int_a^b t^{-\frac{1}{1-\theta}} (1-t)^{-\frac{1}{1-\theta}} dt, \quad 0 \leq a < b \leq 1.$$

Démonstration. En conditionnant par $[J_1 = i, J_2 = j]$, $i, j = 0, 1$, on obtient alors que

$$\begin{aligned}C(u_1, u_2) &= P(\tilde{U}_1 \leq u_1, \tilde{U}_2 \leq u_2) \\ &= (1-\theta)^2 P(U_1^{1-\theta} \leq u_1, U_2^{1-\theta} \leq u_2) \\ &\quad + \theta(1-\theta) P(U_1^{1-\theta} \leq u_1, U_2^{1-\theta} (1-V) \leq u_2) \\ &\quad + \theta(1-\theta) P(U_1^{1-\theta} V \leq u_1, U_2^{1-\theta} \leq u_2) \\ &\quad + \theta^2 P(U_1^{1-\theta} V \leq u_1, U_2^{1-\theta} (1-V) \leq u_2)\end{aligned}$$

Par la suite, on peut directement calculer

$$\begin{aligned}I_1 &= P(U_1^{1-\theta} \leq u_1, U_2^{1-\theta} \leq u_2) = u_1^{\frac{1}{1-\theta}} u_2^{\frac{1}{1-\theta}} \\ I_2 &= P(U_1^{1-\theta} \leq u_1, U_2^{1-\theta} (1-V) \leq u_2) = \frac{1}{\theta} u_1^{\frac{1}{1-\theta}} u_2 - \frac{1-\theta}{\theta} u_1^{\frac{1}{1-\theta}} u_2^{\frac{1}{1-\theta}} \\ I_3 &= P(U_1^{1-\theta} V \leq u_1, U_2^{1-\theta} \leq u_2) = \frac{1}{\theta} u_1 u_2^{\frac{1}{1-\theta}} - \frac{1-\theta}{\theta} u_1^{\frac{1}{1-\theta}} u_2^{\frac{1}{1-\theta}}\end{aligned}$$

Il reste maintenant à calculer le dernier terme de l'équation pour conclure. Pour ce faire, séparons en deux cas distincts : cas 1 : $u_1 \leq 1 - u_2$. Effectuons le calcul pour cette partie.

$$\begin{aligned} & P\left(U_1^{1-\theta}V \leq u_1, U_2^{1-\theta}(1-V) \leq u_2\right) \\ &= \int_0^1 P\left(U_1 \leq \left(\frac{u_1}{t}\right)^{\frac{1}{1-\theta}}, U_2 \leq \left(\frac{u_2}{1-t}\right)^{\frac{1}{1-\theta}}\right) dt = I_4 + I_5 + I_6, \end{aligned}$$

avec

$$\begin{aligned} I_4 &= \int_0^{u_1} P\left(U_2 \leq \left(\frac{u_2}{1-t}\right)^{\frac{1}{1-\theta}}\right) dt = \frac{1-\theta}{\theta} u_2^{\frac{1}{1-\theta}} \left[(1-u_1)^{-\frac{\theta}{1-\theta}} - 1\right], \\ I_5 &= \int_{1-u_2}^1 P\left(U_1 \leq \left(\frac{u_1}{t}\right)^{\frac{1}{1-\theta}}\right) dt = \frac{1-\theta}{\theta} u_1^{\frac{1}{1-\theta}} \left[(1-u_2)^{-\frac{\theta}{1-\theta}} - 1\right], \\ I_6 &= \int_{u_1}^{1-u_2} P\left(U_1 \leq \left(\frac{u_1}{t}\right)^{\frac{1}{1-\theta}}, U_2 \leq \left(\frac{u_2}{1-t}\right)^{\frac{1}{1-\theta}}\right) dt = B_\theta(u_1, 1-u_2). \end{aligned}$$

Maintenant, traitons le second cas : $u_1 > 1 - u_2$. De façon analogue au premier cas, on peut clairement démontrer que

$$\int_0^1 P\left(U_1 \leq \left(\frac{u_1}{t}\right)^{\frac{1}{1-\theta}}, U_2 \leq \left(\frac{u_2}{1-t}\right)^{\frac{1}{1-\theta}}\right) dt = I_7 + I_8 + I_9,$$

d'où

$$\begin{aligned} I_7 &= \int_0^{1-u_2} P\left(U_2 \leq \left(\frac{u_2}{1-t}\right)^{\frac{1}{1-\theta}}\right) dt = \frac{1-\theta}{\theta} \left(u_2 - u_2^{\frac{1}{1-\theta}}\right), \\ I_8 &= \int_{1-u_2}^{u_1} dt = u_1 + u_2 - 1, \\ I_9 &= \int_{u_1}^1 P\left(U_1 \leq \left(\frac{u_1}{t}\right)^{\frac{1}{1-\theta}}\right) dt = \frac{1-\theta}{\theta} \left(u_1 - u_1^{\frac{1}{1-\theta}}\right). \end{aligned}$$

Pour conclure, si $u_1 \leq 1 - u_2$,

$$\begin{aligned} C(u_1, u_2) &= (1-\theta)^2 I_1 + \theta(1-\theta) I_2 + \theta(1-\theta) I_3 + \theta^2 (I_4 + I_5 + I_6) \\ &= (1-\theta) \left\{ u_1^{\frac{1}{1-\theta}} (u_2 - \theta) + u_2^{\frac{1}{1-\theta}} (u_1 - \theta) - (1-\theta) u_1^{\frac{1}{1-\theta}} u_2^{\frac{1}{1-\theta}} \right\} \\ &\quad + \theta(1-\theta) \left[u_2^{\frac{1}{1-\theta}} (1-u_1)^{-\frac{\theta}{1-\theta}} + u_1^{\frac{1}{1-\theta}} (1-u_2)^{-\frac{\theta}{1-\theta}} \right] + \theta^2 B_\theta(u_1, 1-u_2) \\ &= (1-\theta) \left\{ \left(1 - u_1^{\frac{1}{1-\theta}}\right) \left(1 - u_2^{\frac{1}{1-\theta}}\right) + u_1 u_2 \left[1 - \left(1 - u_1^{\frac{\theta}{1-\theta}}\right) \left(1 - u_2^{\frac{\theta}{1-\theta}}\right)\right] \right\} \\ &\quad + \theta(1-\theta) \left[u_2^{\frac{1}{1-\theta}} (1-u_1)^{-\frac{\theta}{1-\theta}} + u_1^{\frac{1}{1-\theta}} (1-u_2)^{-\frac{\theta}{1-\theta}} - 1 \right] + \theta^2 B_\theta(u_1, 1-u_2) \end{aligned}$$

et si $u_1 > 1 - u_2$,

$$\begin{aligned} C(u_1, u_2) &= (1-\theta)^2 I_1 + \theta(1-\theta) I_2 + \theta(1-\theta) I_3 + \theta^2 (I_7 + I_8 + I_9) \\ &= (1-\theta) \left\{ \left(1 - u_1^{\frac{1}{1-\theta}}\right) \left(1 - u_2^{\frac{1}{1-\theta}}\right) + u_1 u_2 \left[1 - \left(1 - u_1^{\frac{\theta}{1-\theta}}\right) \left(1 - u_2^{\frac{\theta}{1-\theta}}\right)\right] \right\} \\ &\quad + \theta(u_1 + u_2 - 1) \end{aligned}$$

ce qui complète la démonstration.

□

5.3.1 Algorithme de simulation de la copule de survie de la famille $\mathcal{E}(\theta, \lambda_1, \lambda_2)$

Tout comme pour la représentation stochastique la copule de survie associée à $\mathcal{E}(\theta, \lambda_1, \lambda_2)$ décrit par le modèle de Raftery, on peut également reproduire la méthode pour le cas de la dépendance négative, permettant de générer facilement des données issues de cette copule. De même, l'algorithme ci-dessous indique les nouvelles étapes principales de simulation de données provenant de cette dernière :

1. Générer des aléas indépendantes u_1, u_2, u de loi uniforme sur $[0, 1]$
2. Générer des aléas j_1 et j_2 de la de bernoulli de paramètre θ
3. Posons $v = u_1^{1-\theta} u^{j_1}$ et $w = u_2^{1-\theta} (1 - u)^{j_2}$
4. Le couple désiré sera (v, w)

Ci-dessous quelques graphiques illustrant le comportement de données générées à partir de la copules survie de la famille $\mathcal{E}(\theta, \lambda_1, \lambda_2)$.

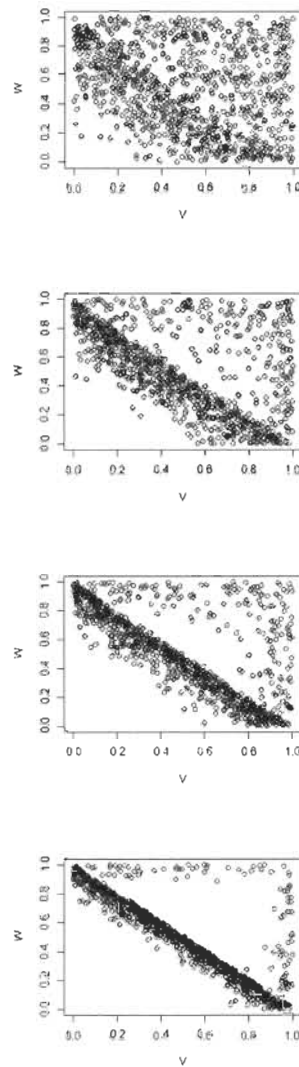


FIGURE 5.1 – Simulation de la copule de Raftery avec $\theta = 0.7, 0.85, 0.9$ et 0.95

5.4 Rho de Spearman

Le calcul de rho the Spearman de ce modèle peut être obtenu rapidement, en exploitant la représentation stochastique du modèle décrite par les équations établies dans (5.3).

Théorème 5.4.1.

$$\rho_C = -\frac{\theta^2}{(2 - \theta)^2}$$

Démonstration. Le rho de Spearman de la copule C représente le coefficient de corrélation associé à la paire aléatoires $(\tilde{U}_1, \tilde{U}_2)$ définie par (5.3). Il en résulte que

$$\rho_C = 12E(\tilde{U}_1\tilde{U}_2) - 3$$

d'où

$$\begin{aligned} E(\tilde{U}_1\tilde{U}_2) &= (1-\theta)^2 E(U_1^{1-\theta}U_2^{1-\theta}) + \theta(1-\theta)E(U_1^{1-\theta}U_2^{1-\theta}V) \\ &\quad + \theta(1-\theta)E(U_1^{1-\theta}U_2^{1-\theta}(1-V)) + \theta^2 E(U_1^{1-\theta}U_2^{1-\theta}V(1-V)) \\ &= \frac{(1-\theta)^2}{(2-\theta)^2} + \frac{\theta(1-\theta)}{(2-\theta)^2} + \frac{\theta^2}{6(2-\theta)^2} \\ &= \frac{\theta^2 - 6\theta + 6}{(2-\theta)^2}. \end{aligned}$$

Le résultat proposé précédemment découle directement des équations précédentes. \square

Dans le cas de cette copule, le calcul du tau de Kendall sera omis, étant donnée la complexité de la forme de la copule.

5.5 Étude de l'estimateur des moments du paramètre de dépendance de la copule de Raftery au cas de dépendance négative

5.5.1 Estimateur du paramètre de dépendance à partir de la corrélation

Proposition 5.5.1. *L'estimateur de moment du paramètre de dépendance θ basé sur la corrélation de Person est donné par*

$$\hat{\theta}_{cp} = \sqrt{\frac{r}{1 - \frac{\pi^2}{6}}}.$$

Démonstration. Rappelons que la structure de corrélation du modèle exponentiel né-

gatif $\mathcal{E}^-(\theta, \lambda_1, \lambda_2)$ est exprimée en fonction de $\theta \in [0, 1]$ et ρ_{\min} . Ainsi nous avons

$$\begin{aligned} r &= \left(1 - \frac{\pi^2}{6}\right) \theta^2 \\ \theta^2 &= \frac{r}{\left(1 - \frac{\pi^2}{6}\right)} \\ \hat{\theta}_{cp} &= \sqrt{\frac{r}{\left(1 - \frac{\pi^2}{6}\right)}}. \end{aligned}$$

Ce qui conduit à un estimateur du paramètre de dépendance θ pour le modèle négatif. Nous observons que la branche positive est uniquement considérée car $\theta \in [0, 1]$. \square

5.5.2 Estimateur du paramètre de dépendance à partir de la corrélation de Spearman

Proposition 5.5.2. *L'estimateur de moment du paramètre de dépendance θ basé sur la corrélation de Spearman est donné par*

$$\hat{\theta}_{sp} = \frac{2\sqrt{-r}}{1 + \sqrt{-r}}$$

Démonstration. Rappelons que la structure de corrélation de Spearman du modèle exponentiel négatif $\mathcal{E}^-(\theta, \lambda_1, \lambda_2)$ est exprimée en fonction de $\theta \in [0, 1]$ uniquement. Ainsi nous avons

$$\begin{aligned} r &= \frac{-\theta^2}{(2 - \theta)^2} \\ -r &= \frac{\theta^2}{(2 - \theta)^2} \\ \sqrt{-r} &= \frac{\theta}{(2 - \theta)} \\ 2\sqrt{-r} &= \theta(1 + \sqrt{-r}) \\ \hat{\theta}_{sp} &= \frac{2\sqrt{-r}}{1 + \sqrt{-r}} \end{aligned}$$

Alors, voici un second estimateur du paramètre de dépendance θ pour le modèle négatif. Nous observons que la branche positive est uniquement considérée car $\theta \in [0, 1]$. \square

5.6 Comparaison des estimateurs par simulation

Pour conclure l'étude du nouveau modèle exponentiel bidimensionnel de type Raftery dans le cas de dépendance négative, effectuons une comparaison de l'étude de l'estimateur des moments du paramètre de dépendance démontré précédemment. L'exercice est le même que celui effectué au chapitre précédent. Effectuons un tableau avec les estimateurs de θ . Pour différentes tailles d'échantillons, on obtient les résultats ci-dessous.

Estimateur	n=20	n=30	n=50
$\hat{\theta}_{sp}$	0.72983	0.81482	0.79198
EQM($\hat{\theta}_{sp}$)	0.02242	0.00505	0.00511
$\hat{\theta}_{cp}$	0.73714	0.85453	0.79241
EQM($\hat{\theta}_{cp}$)	0.05221	0.01662	0.01377

FIGURE 5.2 – Simulation des estimateurs de la copule de Raftery négative avec $\theta = 0.6$

Estimateur	n=100	n=150	n=200
$\hat{\theta}_{sp}$	0.79095	0.81613	0.79444
EQM($\hat{\theta}_{sp}$)	0.00189	0.00107	0.00238
$\hat{\theta}_{cp}$	0.81481	0.86881	0.82778
EQM($\hat{\theta}_{cp}$)	0.00650	0.00614	0.00684

FIGURE 5.3 – Simulation des estimateurs de la copule de Raftery négative avec $\theta = 0.6$

Estimateur	n=20	n=30	n=50
$\hat{\theta}_{sp}$	0.77204	0.76956	0.74187
EQM($\hat{\theta}_{sp}$)	0.02113	0.01965	0.01496
$\hat{\theta}_{cp}$	0.75584	0.81230	0.77445
EQM($\hat{\theta}_{cp}$)	0.08444	0.03254	0.0211

FIGURE 5.4 – Simulation des estimateurs de la copule de Raftery négative avec $\theta = 0.8$

Estimateur	n=100	n=150	n=200
$\hat{\theta}_{sp}$	0.83816	0.79102	0.79103
EQM($\hat{\theta}_{sp}$)	0.00391	0.00469	0.00102
$\hat{\theta}_{cp}$	0.86442	0.8381	0.81835
EQM($\hat{\theta}_{cp}$)	0.00620	0.00511	0.00339

FIGURE 5.5 – Simulation des estimateurs de la copule de Raftery négative avec $\theta = 0.8$

$$EQM(\hat{\theta}_{sp}) = \frac{1}{k} \sum_{i=1}^k (\hat{\theta}_{sp} - \theta)^2$$

$$EQM(\hat{\theta}_{cp}) = \frac{1}{k} \sum_{i=1}^k (\hat{\theta}_{cp} - \theta)^2$$

Tout comme dans le chapitre précédent, pour vérifier la qualité des estimateurs, considérons l'erreur quadratique moyenne des estimateurs simulés pour seulement $k = 5$ répétitions cette fois-ci pour la compilation. Tout comme dans le cas du modèle exponentiel de Raftery dans le cas de dépendance positive, peu importe la taille d'échantillon, l'estimateur du paramètre de dépendance par la corrélation de Person reste imprécis. On favorise l'estimateur par la corrélation de Spearman.

5.6.1 Algorithme de simulation de la fonction de survie de la famille $\mathcal{E}^-(\theta, \lambda_1, \lambda_2)$

Tout comme pour la simulation de données issues de la famille $\mathcal{E}(\theta, \lambda_1, \lambda_2)$, la simulation de données issues de la nouvelle famille $\mathcal{E}^-(\theta, \lambda_1, \lambda_2)$ s'effectue également via la représentation stochastique d'un couple de variables aléatoires $(X, Y) \in \mathcal{E}^-(\theta, \lambda_1, \lambda_2)$ donné par

$$X = (1 - \theta)Z_1 + J_1Z_3,$$

$$Y = (1 - \theta)Z_2 + J_1Z_4,$$

où Z_1, Z_2, Z_3, Z_4 , des variables aléatoires indépendantes telles que $Z_1 \sim \text{Exp}(\lambda_1)$, $Z_2 \sim \text{Exp}(\lambda_2)$, $Z_3 \sim \text{Exp}(\lambda_1)$ et $Z_4 \sim \text{Exp}(\lambda_2)$. Puis J_1 et J_2 qui suit une loi de Bernoulli de paramètre θ indépendante de Z_1, Z_2, Z_3 et Z_4 . Ainsi l'algorithme ci-dessous indique les étapes principales de simulation de données provenant de la famille $\mathcal{E}^-(\theta, \lambda_1, \lambda_2)$:

1. Fixons λ_1, λ_2 et θ
2. Générer des aléas indépendantes u_1, u_2, u de loi uniforme sur $[0, 1]$
3. Posons $z_1 = -\frac{1}{\lambda_1} \ln(u_1)$, $z_2 = -\frac{1}{\lambda_2} \ln(u_2)$, $z_3 = -\frac{1}{\lambda_1} \ln(u)$ et $z_4 = -\frac{1}{\lambda_2} \ln(1 - u)$.
4. Générer des aléas j_1 et j_2 de la loi de bernoulli de paramètre θ
5. Posons

$$x = (1 - \theta)z_1 + j_1z_3, \quad y = (1 - \theta)z_2 + j_2z_4.$$

6. Le couple désiré sera (x, y)

Tout comme pour le modèle précédent, une autre manière de générer des données de la famille $\mathcal{E}(\theta, \lambda_1, \lambda_2)$ consiste à simuler des observations (u, v) de la copule de survie \hat{C} , puis poser $x = \exp(-\lambda_1 u)$ et $y = \exp(-\lambda_2 v)$. Ainsi le couple (x, y) provient de la fonction de survie \bar{H} .

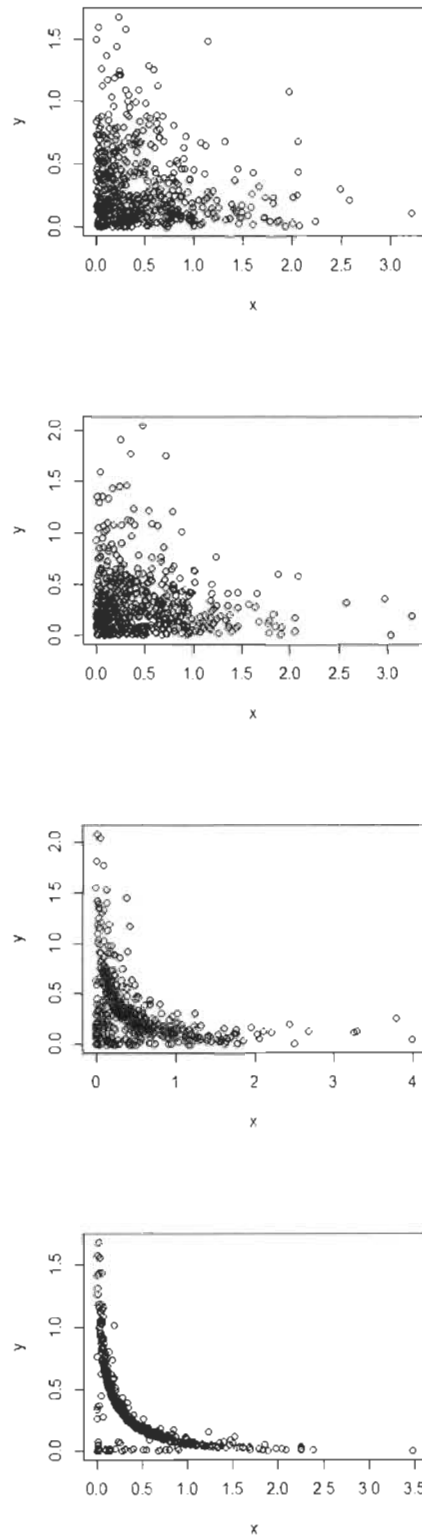


FIGURE 5.6 – Simulation de la fonction de survie du modèle de Raftery négative avec $\theta = 0.3, 0.5, 0.8$ et 0.95

Chapitre 6

Conclusion

Le présent travail de recherche étudie le comportement de la nouvelle loi exponentielle bidimensionnelle de type Raftery dans le cas de la dépendance positive et négative, ainsi que les estimateurs du paramètre de dépendance à l'aide de la méthode de chocs comonotones récemment introduit par Genest, Mesfioui et Shoulz (2018). La contribution originale de ce travail consiste à proposer et étudier deux extensions du modèle classique de Raftery. Ces modèles se basent sur la méthode des chocs comonotones.

Pour faire un retour sur les objectifs principaux de ce travail, l'extension du modèle exponentiel bidimensionnel de Raftery au cas de variables aléatoires non échangeables a été établie. Par ailleurs, le présent travail a permis de constater que pour toutes les simulations des estimateurs du paramètre de dépendance que ce soit pour les deux modèles de Raftery proposés, l'estimateur par la corrélation de Spearman est celle qui possède la plus petite erreur quadratique moyenne. Le présent travail permet de conclure que l'estimation par la corrélation de Pearson est moins précise. Plus spécifiquement, nous avons déduit une mesure de concordance puis la structure de corrélation de ce modèle pour établir un estimateur du paramètre de dépendance. Une étude comparative de ces deux approches basées sur plusieurs simulations a été explorée.

Le présent travail a également permis d'établir des algorithmes de simulation de la fonction de survie des deux familles de Raftery étudiés précédemment. Quelques simulations ont été effectuées pour constater la tendance de la survie du modèle lorsque le paramètre de dépendance augmente. La copule de Raftery a également été simulée pour observer la tendance linéaire du modèle dans le second cas.

Bibliographie

- [1] Genest, C. & MacKay, J. (1986). Copules archimédiennes et familles de lois bidimensionnelles dont les marges sont données. *Canad J Statist* 14 : 145–159.
- [2] Genest, C., Mesfioui, M. & Schulz, J. (2018). A new bivariate Poisson common shock model covering all possible degrees of dependence. *Statistics Probability Letters* 140 : 202–209.
- [3] Nelson, RB.(1992). On measures of association as measures of positive dependence. *Statist Probab Lett* 14 : 269–274.
- [4] Nelson, RB.(1995). Copulas, characterization, correlation, and counterexamples. *Math Mag* 68 : 193–198.
- [5] Nelson, RB.(2006). An introduction to copulas. *Springer Series in Statistics*.
- [6] Raftery, AE (1984). A continuous multivariate exponential distribution. *Comm Statist A-Theory Methods* 13 : 947–965.
- [7] Raftery, AE (1985). Some properties of a new continuous bivariate exponential distribution. *Statist Decisions Supplement Issue 2* : 53–58.
- [8] Ross, SM (2014). Initiation to probabilities models *Library of Congress Cataloging-in-Publication Data*
- [9] Sklar, A.(1959). Fonctions de répartition à n dimensions et leurs marges. *Publ Inst Statist Univ Paris* 68 : 193–198.
- [10] Sklar, A.(1973). Random variables, joint distributions, and copulas. *Kybernetika* 9 : 449–460.